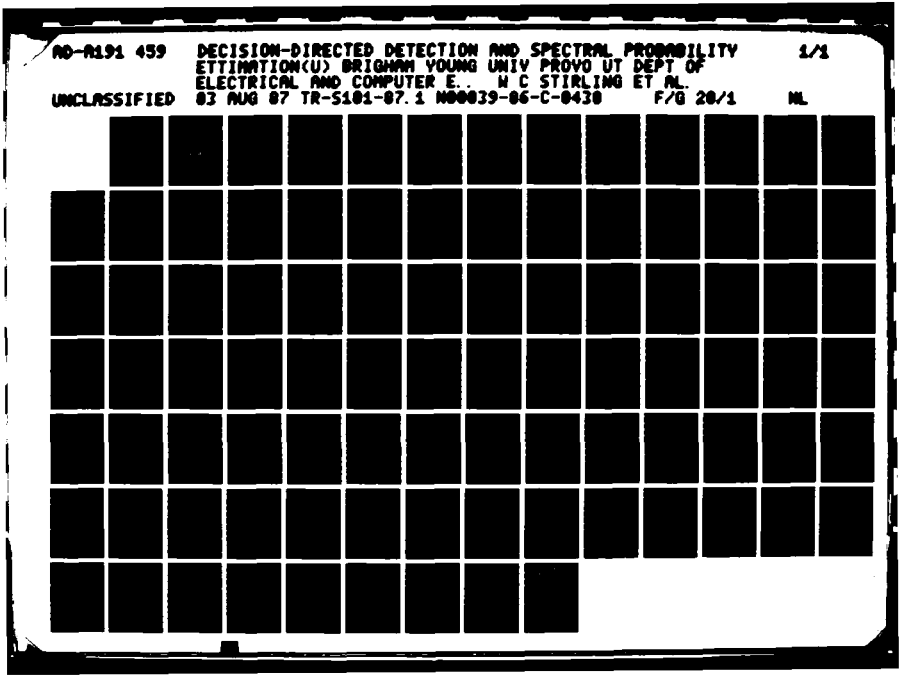
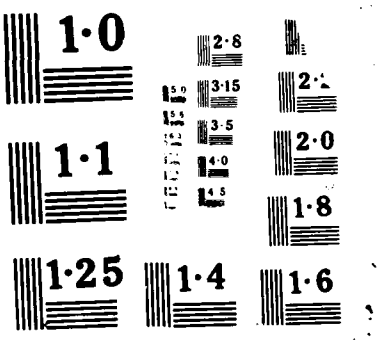

AD-A191 459 DECISION-DIRECTED DETECTION AND SPECTRAL PROBABILITY 1/1
ETTIMATION(U) BRIGHAM YOUNG UNIV PROVO UT DEPT OF
ELECTRICAL AND COMPUTER E.. W C STIRLING ET AL.
UNCLASSIFIED 03 AUG 87 TR-S101-87.1 N00039-86-C-0430 F/G 20/1 NL


1·0
1·1
1·25
1·4
1·6
2·0
1·8

DECISION-DIRECTED DETECTION AND SPECTRAL PROBABILITY ESTIMATION

January 6, 1988

Final Report

Contract No. N00039-86-C-0430

Approved for public release: distribution is unlimited.

COLLEGE OF
ENGINEERING AND TECHNOLOGY
BRIGHAM YOUNG UNIVERSITY
PROVO, UTAH

ELECTRICAL AND COMPUTER ENGINEERING DEPARTMENT
BRIGHAM YOUNG UNIVERSITY · PROVO, UTAH

TR S101-87.1

DECISION-DIRECTED DETECTION AND SPECTRAL PROBABILITY ESTIMATION

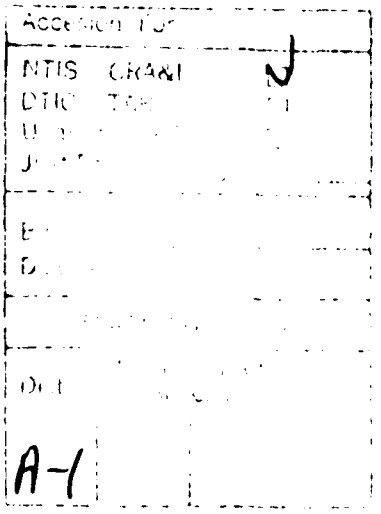
January 6, 1988

Final Report

Contract No. N00039-86-C-0430

W. C. Stirling

R. A. Muir



Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) TR S101-87.1	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Brigham Young University Electrical & Comp. Eng. Dept. — 6b OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of the Cheif of Naval Research
6c. ADDRESS (City, State, and ZIP Code) 459 Clyde Building Provo, Utah 84602	7b. ADDRESS (City, State, and ZIP Code) Arlington, Virginia 22217-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION — 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N 00039-86-C-0430
8c. ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO | PROJECT NO | TASK NO | WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
DECISION-DIRECTED DETECTION AND SPECTRAL PROBABILITY ESTIMATION (U)

12 PERSONAL AUTHOR(S)
Wynn C. Stirling and Robert A. Muir

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM 86 TO 87	August 3, 1987	88

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Decision-Directed, Empirical Bayes, Spectral Probability Estimation

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Over the past several years many contributions have been made to the problem of detecting underwater acoustic signals and estimating signal parameters such as time-of-arrival, frequency-of-arrival, etc. In the previous research effort, the spatial detection problem was re-examined from a decision-directed point of view, and a methodology was presented to estimate the time-varying joint prior spatial distribution of the signal. In the current research effort, attention is directed toward improving detection of temporal signals according to the frequency content of the signal. The source environment characterizes the presence or absence of the various sources (both signals and noise) as they evolve in frequency and time. It may not be assumed that this environment is stationary in either frequency content or time, hence it is necessary to characterize and track the nonstationary behavior of this environment. Knowledge of the source environment may be used to

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL James G. Smith	22b TELEPHONE (Include Area Code) — 22c OFFICE SYMBOL

DD Form 1473, JUN 86 — Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. (continued)

effect on-line adaptation of the decision strategies used to detect hostile targets, and has potential for modifying the collection procedures. The key result of this report is the development of a decision-directed empirical Bayes decision rule which permits a nonstationary prior marginal probability distribution to be estimated (i.e., tracked) based upon the time-varying frequency content of the signal.

Abstract

Over the past several years many contributions have been made to the problem of detecting underwater acoustic signals and estimating signal parameters such as time-of-arrival, frequency-of-arrival, angle-of-arrival, etc. In the previous research effort, the spatial detection problem was re-examined from a decision-directed point of view, and a methodology was presented to estimate the time-varying joint prior spatial distribution of the signal. In the current research effort, attention is directed toward improving detection of temporal signals according to the frequency content of the signal. The source environment characterizes the presence or absence of the various sources (both signals and noise) as they evolve in frequency and time. It may not be assumed that this environment is stationary in either frequency content or time, hence it is necessary to characterize and track the nonstationary behavior of this environment. Knowledge of the source environment may be used to effect on-line adaptation of the decision strategies used to detect hostile targets, and has potential for modifying the collection procedures. The key result of this report is the development of a decision-directed empirical Bayes decision rule which permits a nonstationary prior marginal probability distribuiton to be estimated (i.e., tracked) based upon the time-varying frequency content of the signal.

Contents

1 Introduction **1-1**

1.1 Problem Statement 1-1

1.2 Summary of Technical Approach 1-4

1.3 Summary of Results and Conclusions 1-9

2 Technical Approach **2-1**

2.1 Decision Theory Background 2-1

2.2 Frequency Distribution Estimation 2-7

2.2.1 Probability Models 2-10

2.2.2 Estimation Procedure 2-12

2.2.3 Covariance of Estimation Error 2-15

2.3 Decision-Directed Detection 2-16

2.3.1 Detector Design 2-16

2.3.2 Bias Correction 2-18

3 Performance Evaluation by Monte Carlo Analysis **3-1**

3.1 Signal Modeling Assumptions 3-1

3.1.1 Observations Model 3-2

3.1.2 Markov Chain Model 3-5

3.1.3 Estimation of Signal Parameters 3-6

3.2 Spectral Probability Estimation 3-8

3.3 Monte Carlo Simulation Results 3-8

3.3.1 Time-varying Probability Tracking 3-8

3.3.2 Receiver Operating Characteristics 3-27

Bibliography Bib-1

A Key Results from Martingale Theory A-1

A.1 Background A-1

A.2 Discrete-Time Point Processes A-4

A.2.1 Doob Decomposition A-5

A.2.2 Estimation from Discrete-Time Point Processes A-6

B Envelope-squared Detection B-1

B.1 Envelope of a Narrowband Process B-1

B.2 Envelope Squared of a Sinusoid plus Narrowband Noise B-2

B.3 DFT Bin Noise Correlation B-3

B.4 Detector Structure B-4

C Distribution List C-1

List of Figures

1.1 Signal Processing Flow–Feedback Among Subsystems 1-5

2.1 Feed-Forward Rules 2-3

2.2 Feedback, or Decision-Directed, Rules 2-4

2.3 Bayes Envelope Function 2-5

2.4 Decision-Directed Spectral Probability Estimation Block Diagram 2-20

3.1 True Time-varying *A Priori* Probability Structure. 3-9

3.2 Estimated Time-Varying *A Priori* Probability Structure (SNR=12dB). . . . 3-11

3.3 Monte Carlo Average for the Estimated *A Priori* Probability Structure (SNR=12dB). 3-12

3.4 Monte Carlo Variance on the Estimated *A Priori* Probability Structure (SNR=12dB). 3-13

3.5 Mean-square Error for the Estimated *A Priori* Probability Structure (SNR=12dB).3-14

3.6 Estimated *A Priori* Probability Structure (SNR=6dB). 3-15

3.7 Monte Carlo Averages for the Estimated *A Priori* Probability Structure (SNR=6dB). 3-16

3.8 Monte Carlo Variance on the Estimated *A Priori* Probability Structure (SNR=6dB). 3-17

3.9 Mean-square Error for the Estimated *A Priori* Probability Structure (SNR=6dB).3-18

3.10 Estimated *A Priori* Probability Structure (SNR=12dB). 3-19

3.11 Monte Carlo Averages for the Estimated *A Priori* Probability Structure (SNR=12dB). 3-20

3.12 Monte Carlo Variance on the Estimated *A Priori* Probability Structure (SNR=12dB). 3-21
3.13 Mean-square Error for the Estimated *A Priori* Probability Structure (SNR=12dB).3-22
3.14 Estimated *A Priori* Probability Structure (SNR=6dB). 3-23
3.15 Monte Carlo Averages for the Estimated *A Priori* Probability Structure (SNR=6dB). 3-24
3.16 Monte Carlo Variance on the Estimated *A Priori* Probability Structure (SNR=6dB). 3-25
3.17 Mean-square Error for the Estimated *A Priori* Probability Structure (SNR=6dB).3-26
3.18 Receiver Operating Characteristics - Known Amplitude, Estimated *A Priori* Probabilities. 3-27
3.19 Receiver Operating Characteristics - Estimated Amplitude and *A Priori* Probabilities. 3-28

Chapter 1

Introduction

1.1 Problem Statement

The problem addressed by this investigation is a continuation of the issues addressed by [34], and represents a natural extension of that research. The practical problem which motivates this study is that of detecting, localizing, and classifying foreign submarine activity by means of acoustic surveillance sensors located at strategic positions, particularly in locations where noise sources are present which emit significant energy in the frequency bands of interest (e.g., ice noise). These noise sources often occur as narrowband "tonals" that are difficult to separate from legitimate targets. Consequently, the target detection problem is greatly complicated by their presence, and an important problem is to develop procedures for distinguishing real targets from pseudo targets. Due to the nature of the signals involved, much harmonic information is generated which may allow the system to identify source characteristics. For example, underwater sources radiate narrow- and broad-band acoustic energy due to propulsion systems, auxiliary machinery, and hydrodynamic effects [1]. These effects comprise what is generally referred to as the source *signature*. Since these effects occur at different levels and frequencies depending on the actual shape, size, and operating mode of the vessel, the observed signal components can be compared with a set of known target characteristics for identification purposes. Hence, it is very advantageous to estimate the spectrum of the various received signals for identification purposes. For this research effort, we concentrate not on the usual spectral power or energy estimation methods, but on the application of the above analysis for the estimation

of the *probability* of spectral signal presence as a method for more robust identification procedures.

The major thrust of this investigation is the development of adaptive methods for estimating the signal environment as characterized by the prior probabilities of signal content. These methods lead to the structuring of decision-directed detection procedures that are capable of real-time adaptation to a changing environment (i.e., nonstationary priors).

Review of Previous Research

The investigation conducted under the previous contract (N00039-85-C-0223) [34,35,33] provides encouraging results regarding adaptive classification of signals. That work concentrates primarily on the classification of the spatial frequency content of signals, and yields a decision-directed empirical Bayes methodology to adaptively adjust the decision rule to account for the current joint prior distribution of signals of interest. The resulting detector represents a potential alternative to classical Neyman-Pearson, minimax, and Bayesian decision rules.

The algorithm central to the success of this adaptive decision rule is a recursive, nonlinear, exact minimum mean square estimator proposed by Segall [29] and first applied to decision-directed detection by Stirling [32]. This algorithm was extended to multivariate detection in [34,33].

The spatial detection problem is that of detecting threat signals from spatially separated acoustic surveillance sensors. The signal is often embedded deeply in the noise background, and standard decision criteria yield marginal results. In order to apply the Bayes formula, one must know the *a priori* distribution of the signal with respect to the spatial coordinates of the detection system. Unfortunately, this is rarely the case and, furthermore, even if known, this distribution would likely be non-stationary (i.e., it would evolve in time and space), since the threat environment is subject to change. Errors in the *a priori* distribution may seriously degrade performance of a Bayes detector in this environment. Furthermore, in a weak signal-to-noise environment, a constant probability of false alarm may result in the probability of a missed detection being excessively large, and decision

rules based upon a specified constant false-alarm probability may be inadequate. Also, minimax decision rules are unduly pessimistic in this environment, and may not lead to acceptable performance.

In view of these issues, the empirical Bayesian approach, is explored in [34], and the prior distribution is estimated from the data. Empirical Bayes procedures are well known, [22], but traditionally deal almost exclusively with the stationary case, wherein the prior distribution is constant. For this case there exist asymptotically sub-minimax decision rules that approach the Bayes envelope. Our problem is somewhat more complex, however, since we must allow this distribution to be **time-varying**. Throughout the course of a collection, the target environment is subject to change as sources move through the collection region, enter and exit sensor beams, etc. One may not be able to wait until all the data are received to make decisions. In fact, a real-time decision making capability is necessary and, critically, it must be able to adaptively adjust the structure of the decision rule to ensure that the decisions are being made as accurately as possible. These constraints on the empirical Bayes procedure are severe, and render classical "feed-forward" decision processes inadequate to deal with the tracking capability. An alternative "feedback" approach is developed in this analysis. Such a *decision-directed* approach represents a significant departure from classical empirical Bayes procedures, but fits well into the general class of adaptive detection procedures such as generalized likelihood ratio tests.

Emphasis of Current Research

The above discussion for spatial frequency has a direct analog for temporal frequency, where the decision problem consists of determining temporal frequency structure rather than spatial frequency structure. In this case also, the empirical Bayes approach is applicable, and may be used to estimate the prior distribution on the frequency content of the signal environment.

An additional issue may be addressed in the temporal frequency context. Consider a collection scenario in an environment where the signal is corrupted by "tonal" noise, such as that encountered in an environment where the movement of ice generates noise tonals that are within the spectrum occupied by threat emitters. In such situations, a key problem

is that of classifying the signal as a threat or as benign (i.e., of natural origin). Thus, the problem becomes one of not only detecting the signal, but of recognizing structural characteristics that serve to separate threat and benign signals.

To address this problem, a robust method of spectral estimation is proposed and examined. This method employs a decision-directed empirical Bayes decision rule to estimate the time-varying prior probability of signal occurrence in a given FFT (fast Fourier transform) bin of a signal. The *a priori* probability of spectral content for each bin is modeled as a finite-state Markov chain and the state of this chain is estimated by obtaining the Doob decomposition of the discrete-time point process representing the decisions. A generalized empirical Bayes likelihood ratio test is used as the detector which feeds decisions to the state estimator.

The present research is focused upon the application of decision-directed empirical Bayes methods to estimate the probability of spectral energy independently for discrete frequencies. This approach does not employ information about any harmonic dependencies which a practical signal would reasonably be expected to possess. Therefore, potentially more robust detection schemes would take advantage of the harmonic content of the signal to improve the detector performance. One approach which seems viable is to use conditional factorization of the joint probability density function in a distributed-network technique similar to that done by Stirling and Swindlehurst [34]. Another approach which may achieve improved performance is to provide for feedback of post-detection classification decision information to sensitize the harmonically related scalar detectors.

1.2 Summary of Technical Approach

Decision-Directed Analysis

The philosophy of decision-directed procedures is illustrated in Figure 1.1. In this figure, the outputs of the signal processing (including detection) portion of this diagram may be used to generate an estimate of the signal model (including both deterministic and probabilistic aspects of the model) and feed it back into the spatio-temporal signal processing block to modify the structure of the collection/detection system. There are a number of

possible feedback strategies that may be adopted, including the following:

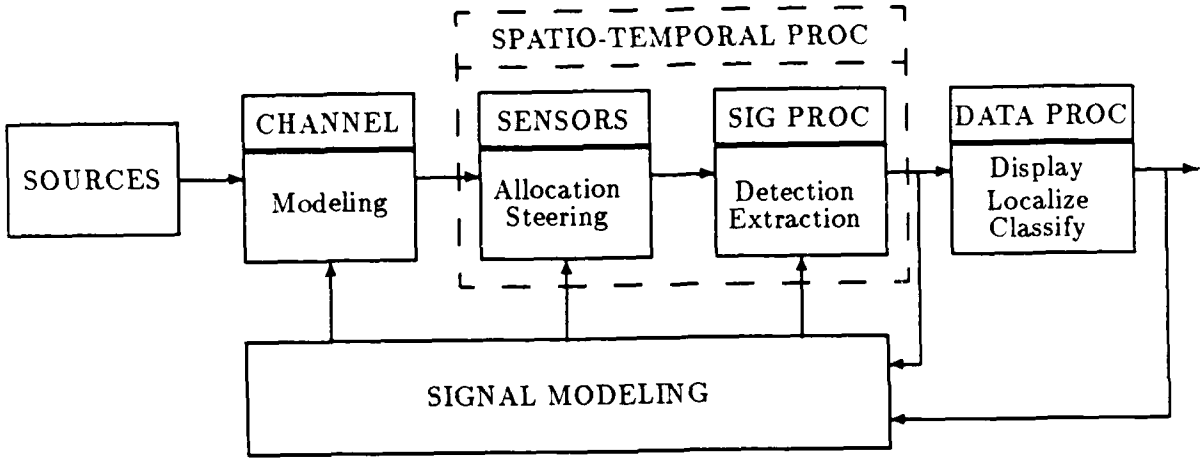


Figure 1.1: Signal Processing Flow–Feedback Among Subsystems

1. *Empirical Bayes.* The most obvious feedback strategy is to modify the decision rules in the empirical Bayes sense. Such a strategy provides estimates of the prior distribution, based upon the past decisions, which modify future behavior of the decision rule. Key issues associated with this procedure are the stability and robustness of the estimator and the ability to track time-varying priors. These issues will be elucidated throughout this report and a multivariate empirical Bayes detection procedure will be introduced.

2. *Control.* A second type of feedback is to use the detected signals and associated estimated probabilities to actually control or modify the spatio-temporal structure of the sensor arrays. For example, array allocation strategies, beamsteering, computer resource allocation, etc., are all possible control strategies which modify or tune the system to improve performance. Technical possibilities will be explored but not fully developed in this report.

3. *Modeling.* A third type of feedback is to use the detected and estimated signal structure to modify the channel model to perform signal enhancement (e.g., channel

equalization) to shape the received signal for more reliable detection and classification. Such techniques are not being considered under this current effort.

The Empirical Bayes Approach

The espousal of the Bayesian approach implies that the unknown state of nature is described by a prior probability distribution. The empirical Bayes decision problem is formulated exactly the same way as a standard Bayes problem, except that prior is unknown and must be estimated from the available data. Suppose that a particular decision problem occurs repeatedly and independently, with the same unknown prior distribution throughout the experiment. (Such a situation obtains when we attempt to detect a weak or fading target in a noisy environment or in the presence of tonal noise). Under this supposition, it is logical to perform analysis on the observation in an attempt to discover the prior distribution. We may define an empirical decision procedure to be a sequence of decision rules which learn or adapt from previous experiments and "converge" in some sense to the true prior. Robbins and related researchers [21,22,24,15] describe the theory of asymptotically optimal decision procedures and demonstrate that such procedures converge to the Bayes envelope function as the number of experiments increases. These results are based upon the assumption that the prior distribution is constant throughout the experiment.

A key aspect of this study is that the prior distribution is not only unknown, it is subject to change as well. This change in the assumptions about the prior constitutes a significant difficulty, since the classical convergence results may no longer be valid. In the extreme case where the changes are completely unpredictable, it is likely true that the empirical Bayes approach is doomed to failure, and some other decision criteria should be evoked. If the changes can be modeled, however, then there is hope that the prior distribution may be "tracked" in a manner entirely analogous to the way a moving target is tracked using, say, a Kalman filter. The key assumption, therefore, is the model that is used to describe the evolution of the distribution. In this study we develop such a model, cast in a state-space environment (analogous to a differential system) and present a recursive state estimator (analogous to a Kalman filter) to estimate the time-varying prior distribution.

To accomplish the goal of tracking the changing distribution, we evoke a particular type

of empirical Bayes decision rule, which may be termed a feedback, or decision-directed rule. Such a rule uses past decisions to estimate the prior, rather than past data directly. To illustrate the difference between a feedback and a feed-forward rule, consider the following simple decision problem: Suppose we observe a signal $y(t)$, for $t = 0, 1, 2, \ldots$, and at each sample, the signal may be white noise with mean zero or with mean $b \neq 0$. The decision problem may be formulated as a hypothesis testing problem of the form

$$\begin{array}{ll} H_0: & y(t) = n(t) \\ H_1: & y(t) = b + n(t) \end{array}, \qquad t = 0, 1, 2, \ldots,$$

where, say, $n(t)$ is an white Gaussian noise sequence of known variance, and b is a known constant. If the prior distribution is known, the problem admits the well-known Bayes solution. If not, the empirical Bayes approach is to estimate this prior. Suppose we consider a feed-forward approach. Let $\pi(t) = P\{H_1 \text{ true at time } t\}$ If $\pi(t)$ is a constant, they we may form an estimate at time t as

$$\hat{\pi}_t = \frac{1}{tb} \sum_{i=1}^{t} y(i)$$

which will converge to the true value as $t \rightarrow \infty$. If $\pi(t)$ is not constant, however, such simple procedures are inadequate. If we pursue the standard empirical Bayes approach, we must postulate an equation of evolution for $\pi(t)$, say $\pi(t+1) = f[\pi(t), t] + w(t)$, a stochastic difference equation. We are also required to ensure that $0 \leq \pi(t) \leq 1$ for all t. Once this model is in place, an appropriate estimation rule must be developed, which will in general be nonlinear.

Alternatively, we may pursue a feedback approach, and deal not with the past values of $y(t)$, but with past decisions. We may model the sequence of past decisions as a *point process*, $\{N\}$, where $N(t) = 1$ if the hypothesis H_1 is selected, and $N(t) = 0$ otherwise. Dealing with $\{N\}$ is much simpler than dealing with the original sequence $\{y\}$. As we shall illustrate in this study, it is indeed possible to formulate physically meaningful procedures to describe the evolution of the probability structure of $N(t)$, which will lead to an estimate of the prior distribution. The point process approach has the great advantage that it leads to a recursive estimator that is easily implemented and can be analyzed theoretically.

Decision-directed detection has been used in many contexts [20,28,17,8,14,10,16] which

deal with the problem of simultaneously detecting a signal and estimating its parameters using decision-directed schemes. Other researchers [6,26,12,27,13] provide analyses of decision-directed procedures for estimating the prior distribution, i.e., decision-directed empirical Bayes procedures. In [30,32] nonstationary empirical Bayes procedures were first introduced, and these were combined with signal parameter estimation in [30,31].

Control Strategies

The signal processing information flow illustrated in Figure 1.1 provides the capability of feeding back signal structure information to the sensor arrays as well as to the detection block. With the empirical Bayes approach, only the prior distribution of the target occurrences is adapted, but the statistical description of the signal itself or the noise is not modified. Furthermore, the allocation of the sensors is not adjusted in any way to accommodate either the probability structure of the signal environment (i.e., the occurrences of threat signals) or of the structure of the signal and noise. Thus, in addition to empirical Bayes decision strategies, there arises the potential for spatio-temporal adaptive control of the collectors and of the detector, including the following:

- *Signal extraction from ice noise.* In collection environments where the noise environment contains impulsive noise, the probability distribution of the signal with respect to its frequency content may be used to discriminate between threat sources and benign (i.e., natural) sources.

- *Mixture process modeling.* Impulsive noise fields may be modeled as mixture processes, wherein the noise process is distributed as a convex linear combination of, say, Gaussian noise. Decision-directed procedures may be used for estimating the mixture parameter.

- *Resource allocation.* Knowledge of the multivariate signal occurrence probability is potentially valuable in making decisions concerning the use of available collection resources. For example, it may be possible to perform adaptive beamsteering to cover regions of high-threat probability more thoroughly, and thereby positively establish

the threat status of the signal. Additionally, this knowledge may motivate a decision to employ additional collection resources that may be held in reserve (due, for example, to power limitations) for high-threat situations.

- *Feedback of classification decisions.* Once sources have been classified as to their frequency content or signature, this information can be supplied to the detectors to improve detection performance in the harmonically related bins represented in the classification decision. Tonal noise which is nonharmonic would be classified as such and those classes without harmonic structure could be given a lower weighting in sensitizing the detectors.

1.3 Summary of Results and Conclusions

Technical Results

The results of this study include:

1. *The formulation of a decision-directed empirical Bayes detection strategy for adaptively tuning the decision-rule to match the observed characteristics of the signal environment.* This decision rule results in a generalized likelihood ratio test wherein the *a priori* distribution is modeled as a finite-state Markov chain that is estimated or tracked as a function of past decisions.

2. *The develpoment of a new algorithm to perform* **spectral probability estimation** *via a bank of scalar decision-directed empirical Bayes detectors and estimators.* This algorithm has application in a signal acquisition and analysis scenario when the probability of signal presence at certain frequencies is a critical surveillance factor. In many sonar applications, the spectral content of a signal is the prime characteristic employed to discriminate between threat sources and friendly sources.

3. *The joint estimation of signal distribution parameters simultaneously with the estimation of the prior distribution.* The empirical Bayes approach requires the estimation of the prior distribution, but, classically, assumes that the conditional distributions

are known. This analysis provides an algorithm for estimating the signal magnitude and noise variance as well.

Conclusions

The conclusions of this study are

1. Decision-directed empirical Bayes procedures in spectral probability estimation have been shown to be useful in establishing the probability of signal presence at given discrete frequencies. Using simulated data, a number of test scenarios have been conducted and the detector performance has been evaluated. These simulations have shown that effective estimates of the signal probability spectrum are obtained at various signal-to-noise ratios. These results have been presented as probability surfaces plotted against time and frequency. The problem of runaway is demonstrated when signal parameters are also required to be estimated, but is negligible for the case when the prior probability only is estimated. Averaged probability estimates obtained through Monte Carlo analysis demonstrate high confidence in the resulting probability information.

2. The decision-directed rule effectively tracks time-varying prior distributions. A number of time-varying prior distributions have been simulated and it has been shown that these rates may be effectively tracked with a decision-directed approach. The most obvious feature of these estimated rates is a time-lag which is a feature typical of real-time processing. The estimator is causal and, consequently, a few samples are required to lock on to the new rate. This lagging phenomenon cannot be eliminated with real-time processing; only noncausal processing involving a smoothing algorithm is capable of removing such phenomena.

3. The performance of the decision-directed empirical Bayes detector is compared to the standard Bayes case, where the prior is exactly known. It is shown that as the signal-to-noise ratio increases, the performance as measured by the total probability of error for the empirical Bayes approach approaches that of the standard case. A striking aspect of the simulations studied is that the additional complication of estimating

the signal parameters (the signal strength in the cases studied) does not significantly change the performance of the detector.

Chapter 2

Technical Approach

2.1 Decision Theory Background

In order to estimate the *a priori* probabilities associated with the various frequency components of a given signal, we assume the following model: Let $y(t)$ be a received signal composed of possibly harmonically related sinusoids contaminated with additive white Gaussian noise. Thus we can represent $y(t)$ as summation of sinusoids and noise as follows:

$$y(t) = \sum_{i=1}^{P} a_i(t) A_i(t) \cos(\omega_i t + \phi_i) + \nu(t). \tag{2.1}$$

Where ω_i is the frequency of signal i, ϕ_i is the phase angle for signal i, uniformly distributed from 0 to 2π, $A_i(t)$ is the time-varying amplitude of signal i, and $\nu(t)$ is additive Gaussian white noise. The number of sinusoids present is denoted by P, where P has no upper bound. The $a_i(t)$ are discrete-time point processes (DTPP's) describing the signal presence at time t. In estimating the *a priori* probability of the signal presence, we are then actually estimating the rate (i.e. the expected value) of the DTPP governing the signal presence at a given frequency. This model is therefore comprehensive enough to admit all forms of compositions of sinusiods at given frequencies when there is noise present. This model could also be used to investigate other phenomena such as frequency-hop spread-spectrum communication systems and frequency-multiplexed communication channels.

More generally, we can describe the state of nature as the set of hypotheses defined by:

$$\mathbf{H}_{\omega_1,\omega_2,\ldots,\omega_P} : S_1^{\alpha_1} \cap S_2^{\alpha_2} \cap \cdots \cap S_P^{\alpha_P}$$

where S_j is the event that a signal is present at at frequency ω_j. The α_j are binary-valued

indices representative of the logical operation denoting signal presence at ω_j defined as follows:

$$S_j^{\alpha_j} \equiv \begin{cases} S_j & if \quad \alpha_j = 1 \\ \overline{S_j} & if \quad \alpha_j = 0 \end{cases} \tag{2.2}$$

where $\overline{S_j}$ is the logical complement of S_j. Let the intersection of all the $S_j^{\alpha_j}$ be denoted a superclass set and let it be written as

$$S_1^{\alpha_1} \cap S_2^{\alpha_2} \cap \cdots \cap S_P^{\alpha_P} = \mathcal{S}^{\alpha_1 \cdots \alpha_P}. \tag{2.3}$$

Thus, the model given in Equation 2.1 yields observations drawn from the set of 2^P mutually exclusive superclass variables. Thus, the observations $y(t)$ may belong to any of the superclasses at a given time t, and furthermore, the superclass assignment characterizing the observations may vary with time.

The problem scenario may be viewed in the classical framework of detection theory. We will assume that there are P sinusoids present and that there is little or no reliable *a priori* information concerning the probability distribution of the signal classification. We will not assume, however, that the signal classifications are independent.

In order to make subsequent discussion more lucid, at this point we introduce notational conventions to be employed in the remainder of the report. Since the problem is concerned with estimation of frequency content we use a discrete Fourier (DFT) transform to obtain frequency content information about the time-domain observations $y(t)$. Justification of this approach is given later. Let us denote the transformed observations in a given DFT bin as $Y_k(t_\ell)$ where k is the bin index. Due to the fact that the DFT information is given not only for discrete time t, but for block-quantized time, we use the ℓ subscript to indicate the block index at which analysis is being performed. Therefore, we shall refer to the observations from now on as $Y_k(t_\ell)$ being the data from the kth DFT bin taken at time t_ℓ. These observations can be expressed in vector form as $\mathbf{Y}(t_\ell)$.

We proceed with an empirical Bayes approach, and estimate sequentially the *a priori* distribution of the signal for use in an M-ary decision problem. We follow the philosophy espoused by [22] that there exists an unknown *a priori* distribution on the signal structure, and this distribution may be discovered by processing the observations, $\mathbf{Y}(t_\ell)$, over time.

The decision function, $D(\mathbf{Y})$, then, is adaptive, in the sense that $D(\cdot)$ will depend on past observations, i.e.,

$$D[\mathbf{Y}(t_\ell)] = D[\mathbf{Y}(s), s < t_\ell; t_\ell, \mathbf{Y}(t_\ell)] \tag{2.4}$$

such that action $D[\mathbf{Y}(t_\ell)]$ is a function of all past observations, $\mathbf{Y}(s), s < t_\ell$. This approach may be altered somewhat by rendering the decision function as a function of past decisions (which is a function of past observations) rather than as a function of past observations directly. Such a decision-directed rule has the form

$$D[\mathbf{Y}(t_\ell)] = D\{D[\mathbf{Y}(s)], s < t_\ell; t_\ell, \mathbf{Y}(t_\ell)\}. \tag{2.5}$$

Decision rules of the form (2.4) may be termed "feed-forward" decision rules, with information flow as depicted in Figure 2.1, since the past data are fed into a decision function generator which in turn modifies the decision rule. Decision rules of the form (2.5) may be

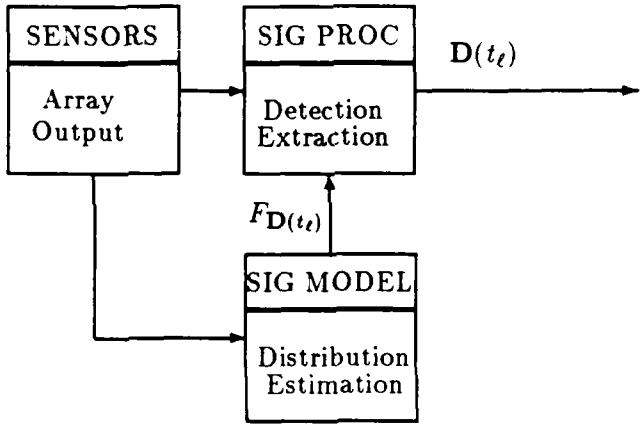


Figure 2.1: Feed-Forward Rules

termed "feedback" decision rules, with information flow as depicted in Figure 2.2, since the past decisions are fed back into a decision function generator which modifies the decision rule. Clearly, a more general decision rule may be formulated which employs both feedback and feed-forward information flows, which represents an obvious generalization of Figures 2.1 and 2.2.

The following discussion is a general representation and uses the time-domain observations $y(t)$, but similar constructions exist for transformed signals of any type. Adaptive decision rules of the types described above may be used to improve performance over non-adaptive procedures, since they may be used to estimate the *a priori* distribution. To

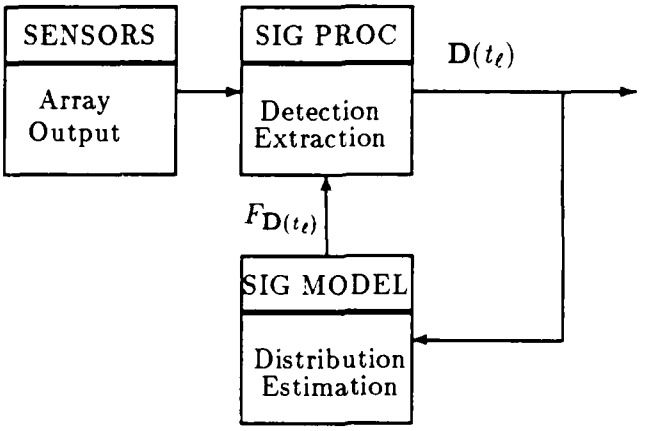


Figure 2.2: Feedback, or Decision-Directed, Rules

illustrate, consider the binary hypothesis problem

$$\begin{aligned} H_0: &\quad y(t) = n(t) \\ H_1: &\quad y(t) = s(t) + n(t) \end{aligned}$$

where $s(t)$ is known and $n(t)$ is of known distribution. If the *a priori* distribution of the occurrence of $s(t)$ were known, the optimal decision rule (in the sense of minimizing the probability of error) would be of the form

$$D[y(t)] = \begin{cases} 1 & \text{if } \pi f(y(t)|H_1) > (\quad - \pi) f(y(t)|H_0) \\ 0 & \text{otherwise} \end{cases}$$

where π is the *a priori* probability of signal occurrence and $f(y(t)|H_1)$ and $f(y(t)|H_0)$ are the probability distributions of $y(t)$ under hypotheses H_1 and H_0, respectively. The problem we face is that π is not known. The danger in arbitrarily guessing the value of π is well known, but is illustrated here for completeness. Let $R(D, \pi)$ denote the Bayes risk function under a distribution for π. The Bayes decision rule will be one that minimizes this function. Let $D_\pi(\cdot)$ denote such a rule. The Bayes envelope function, $r(\pi) = R(D_\pi, \pi)$ represents the minimum risk envelope when the decision rule is specified with the correct value of the *a priori* distribution. For the binary decision rule in this example, the Bayes envelope function is displayed in Figure 2.3. Now it can be seen what happens when the prior is in error. Suppose the true *a priori* distribution is π^* and the assumed *a priori* distribution is $\pi' \neq \pi^*$. Clearly, $D_{\pi'}$ can lead to excessively large risks, and it would perhaps be prudent to employ a minimax rule (denoted by π_M in the figure) which bounds the risk for all values of π. Robbins [22] showed, in a classical result, that empirical Bayes rules

can achieve asymptotically the Bayes envelope, and proposed feed-forward type decision rules of the form

$$D[y(s), s < t; y(t)] = \begin{cases} 1 & \text{if } \hat{\pi}(t) f(y(t)|H_1) > (1 - \hat{\pi}(t)) f(y(t)|H_0) \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{\pi}(t)$ is an estimate of π given $\{y(s), s < t\}$.

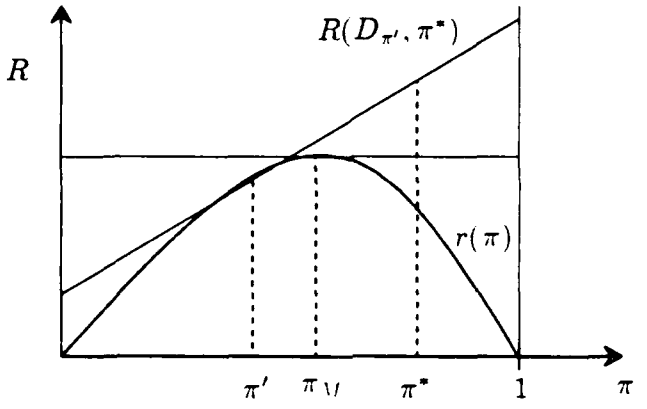


Figure 2.3: Bay s Envelope Function

The use of feedback decision rules for estimation of the prior is perhaps first treated analytically [1] by Davisson and Schwartz [6], wherein decision feedback algorithms are proposed and runaway (a divergence phenomenon which may occur if a sequence of detection errors cause the estimate of the prior to converge to zero or unity, thereby causing the decision rule to go unstable) probabilities are bounded. The resulting decision rule is of the form

$$D[D(s), s < t; y(t)] = \begin{cases} 1 & \text{if } \hat{\pi}(t) f(y(t)|H_1) > (1 - \hat{\pi}(t)) f(y(t)|H_0) \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{\pi}(t)$ is an estimate of π given $\{D_s, s < t\}$.

Although a complete discussion of feedback and feed-forward decision rules will not be attempted, it may be instructive to comment briefly on some of the differences.

[1] Earlier researchers [20,28,17] successfully used decision-directed detectors, but described performance empirically.

1. The feed-forward decision rules are asymptotically sub-minimax. It can be shown that, under appropriate technical conditions [24], the risk using an unbiased estimator of π will converge, asymptotically, to the Bayes envelope.

   No such global results are available for the feedback decision rules, due to the positive probability of runaway. It can be shown, however, that the probability of runaway can be bounded, and these bounds are double exponentially tight [6]. Consequently, barring runaway, the risk using an unbiased feedback estimator of π will converge to the Bayes envelope.

2. Classical, or feed-forward empirical Bayes rules are quite complex, whereas decision-directed rules are extremely simple. Consequently, they are more attractive for use and are more easily implemented.

3. The feed-forward rules are based on the assumption that the *a priori* distribution is stationary (time-invariant). Indeed, stationarity is the very basis of the classical empirical Bayes approach. As noted by [6], many applications where the prior is unknown are highly likely to be nonstationary, and it will be necessary to "track" the nonstationary prior. In such situations, the feed-forward rules may be intractable, and the use of feedback rules may be the only viable approach.

The objective of this investigation is to estimate the multivariate probability distribution function of the signal classification, i.e., the probability, at each time t, that the signal is in any of the possible classification states. We will assume that the distribution function may be time-varying and, therefore, we are required to obtain equations of evolution for this multivariate distribution. The observations that are available for estimating this distribution are the outputs of the sensors. We shall employ a feedback approach, and estimate the distribution based upon the past decisions, thereby developing a rule of the form expressed by (2.5), wherein the multivariate probability distribution function of the detection events is estimated and used to formulate the empirical Bayes decision rule.

2.2 Frequency Distribution Estimation

Initially, in order to estimate the probability distribution of the frequencies present in the signal $y(t)$, we must transform the time-domain data to the frequency domain or do some other form of spectral analysis to establish the energy content of the signal at a given time. We will not attempt a complete exposition of spectral estimation methods in this report, but for the sake of completeness, we illustrate the problem by giving the following examples:

- Short-time Fourier analysis [11] is performed by computing a running average of discrete Fourier transforms of the input signal. This method is useful because of the computational advantages involved in using the Fast Fourier Transform (FFT) algorithm to compute the transformed data and the simplicity of the averaging process. An implicit assumption made when employing Fourier analysis is that the signal is inherently stationary – an assumption which cannot be completely valid for any physically realizable situation, but is, in fact, approximately satisfied for many signals, especially during short time intervals.

- The Wigner distribution [5,4,3,9,19] has recently gained much popularity as a tool in spectral analysis due to the fact that it includes the time parameter in it's formulation. It involves the computation of signal energy at discrete frequencies and therefore is limited in resolution, but the frequency resolution is not limited to the number of time samples as is the case with the discrete Fourier methods. The FFT algorithm can be used to speed up computation of the Wigner distribution, but the computational burden is still greater than the short-time Fourier analysis methods.

- A plethora of so-called high-resolution estimation techniques exist which are not limited to the resolution of energy at discrete frequencies, but are able to resolve energy in a continuous spectrum. Although many variations exist, these methods largely rely on an eigenvalue-eigenvector decomposition of the correlation or covariance matrix generated from the signal data. Some of these methods are more computationally burdensome than others, but, as a class, they do involve a greater computational cost

than either of the above two methods. Some of the more well known high-resolution methods are the MUSIC algorithm [25], a variety of modifications of this approach [2], and ARMA (auto-regressive, moving average) statistical signal models [23].

For our purposes, we choose the discrete Fourier transform method as an appropriate procedure to examine the spectral energy present during a given time interval which we may designate as a block of data samples. During the time block then, we transform the input time data sequence $y(t)$ into a frequency data sequence designated as $Y_k(t_\ell)$. We can alternatively represent this sequence as a frequency data vector $\mathbf{Y}(t_\ell)$ for the block at time t_ℓ.

Let $\mathcal{Y} \in \mathbf{C}^P$ denote the signal space, where $\mathbf{C}$ is the complex plane. And let $\mathbf{Y}(t_\ell) \in \mathcal{Y}$ denote the observed frequency data at time block t_ℓ, for $t_\ell = 0, 1, \ldots$. Let S denote a classification set for the signal $\mathbf{Y}$, and define a decision function

$$D : \mathcal{Y} \mapsto \{0, 1\}$$

as a binary-valued function mapping the observed signal into a abstract classification space. That is,

$$D[\mathbf{Y}(t_\ell)] = \begin{cases} 1 & S \text{ occurs at time } t_\ell \\ 0 & S \text{ does not occur at time } t_\ell \end{cases}$$

For the P classes S_j, $j = 1, \ldots, P$ given above, we may define

$$D_j[\mathbf{Y}(t_\ell)] = \begin{cases} 1 & S_j \text{ occurs at time } t_\ell \\ 0 & S_j \text{ does not occur at time } t_\ell \end{cases} \tag{2.6}$$

and consider the P-vector decision function

$$\mathbf{D}(\cdot) = [D_1(\cdot) \cdots D_P(\cdot)]^T$$

We note that the classification sets S_j need not be disjoint (i.e., $\mathbf{Y}(t_\ell)$ may belong to any or all of the classes). We will assume that $\mathbf{Y}(t_\ell)$ must be classified into at least one of the S_j. Alternatively, using the superclass notation given above, the decision problem reduces to selecting the one superclass possessing the proper attributes.

At each time, t_ℓ, we are confronted with an M-ary decision problem involving the hypotheses

$$H_{\alpha_1 \cdots \alpha_P} : \mathcal{S}^{\alpha_1 \cdots \alpha_P}, \qquad \alpha_1 \cdots \alpha_P \in \{0, 1\}^P$$

where $\mathcal{S}^{\alpha_1 \cdots \alpha_P}$ is defined in Equation 2.3.

The Bayesian approach to this problem is to choose the hypothesis H_j for which the likelihood function

$$\pi^{\alpha_1 \cdots \alpha_P}(t_\ell) f(\mathbf{Y}(t_\ell) | \mathcal{S}^{\alpha_1 \cdots \alpha_P}), \tag{2.7}$$

is maximized, where $f(\mathbf{Y}(t_\ell) | \mathcal{S}^{\alpha_1 \cdots \alpha_P})$ is the distribution of $\mathbf{Y}(t_\ell)$ under classification $\mathcal{S}^{\alpha_1 \cdots \alpha_P}$ at time t_ℓ, and

$$\pi^{\alpha_1 \cdots \alpha_P}(t_\ell) = P\{\mathcal{S}^{\alpha_1 \cdots \alpha_P}; t_\ell\} \tag{2.8}$$

is the *a priori* probability mass function for each of the events $\mathcal{S}^{\alpha_1 \cdots \alpha_P}$ at time t_ℓ. The decision-directed empirical Bayes approach is to estimate $\pi^{\alpha_1 \cdots \alpha_P}(t_\ell)$ by means of the past decisions, $\{\mathbf{D}(s), s < t_\ell\}$.

Ideally, we would wish to estimate $\pi^{\alpha_1 \cdots \alpha_P}(t_\ell)$ directly in order to exploit any information contained in the harmonic relationships which are to be expected in the analysis of complex waveforms. However, as an initial simplification of the problem, we address the case where it is assumed that we can factor $\pi^{\alpha_1 \cdots \alpha_P}(t_\ell)$ into independent marginal probabilities as follows:

$$\pi^{\alpha_1 \cdots \alpha_P}(t_\ell) = \pi^{\alpha_1} \pi^{\alpha_2} \cdots \pi^{\alpha_P}.$$

We later address methods which treat the more general case involving interdependence between the various frequency components. Thus, these marginal probabilities for the various signal classes are the probabilities to be estimated. This model gives rise to a set of scalar estimators each of which provides estimates for a given marginal probability, and is an attractive approach to the problem since the dimensionality of the problem remains relatively low compared with the high dimensionality one would encounter when estimating even a portion of the joint probability structure. The fact that the noise in each frequency bin is uncorrelated with other bins also lends support to this approach (see Appendix B.3). However, a significant amount of information about inter-signal correlation, if there is any, is automatically lost in this process. Alternatives to be explored in mitigating this problem are a joint probability conditional factorization similar to that done by Stirling and Swindlehurst [34], and a new method using classification decision feedback initiated for this research, but not yet fully developed as a theoretically robust strategy.

2.2.1 Probability Models

As a practical matter, even though the signal model admits P different sinusoids in the received signal, when using the discrete Fourier transform (DFT) as our spectral analysis tool, the frequency resolution is dependent on the number of time points taken in the analysis window of $y(t)$. Let the number of time points in the analysis window be denoted by M. Due to the periodicity assumption inherent in the frequency domain, only $\frac{M}{2}+1$ unique frequencies are represented by the elements of $\mathbf{Y}(t_\ell)$. Therefore, the dimensionality of the problem is reduced if $\frac{M}{2}+1 < P$ or increased if $\frac{M}{2}+1 > P$. Since we admit that P may, in fact, be a very large number the chance of dimensionality reduction is actually quite small. In any case however, it is not guaranteed that the P different sinusoids are such that they fall into distinct bins of the DFT. This problem can be alleviated by increasing the size of the analysis window, but the sinusoids still cannot be ensured to be in completely separate bins. Thus, we consider that energy present in a given DFT bin is due to a single sinusoid at a frequency contained in the frequency range of that bin.

Now, define $N_k(t_\ell)$ to be the DTPP generated by the decision process operating on the kth component of the observation vector at time t_ℓ. In other words, denote

$$N_k(t_\ell) = D_k[Y_k(t_\ell)],\ k = 0,\ldots,\frac{M}{2}\ \ t_\ell = \ell M/2,\quad \ell = 0,1,\ldots \tag{2.9}$$

where the usage of $\frac{M}{2}$ is explained above. With this understood, and for the sake of simpler notation, we now designate $L = \frac{M}{2}$. Let $\mathcal{B}_{t_{\ell-1}}$ denote a σ-field generated by all of the factors that may affect the distribution of the process $N(t_\ell)$ at time t_ℓ, and define the marginal conditional probability mass function

$$p_{N_k(t_\ell)}(\alpha_k|\mathcal{B}_{t_{\ell-1}}) = P\{N_k(t_\ell) = \alpha_k|\mathcal{B}_{t_{\ell-1}}\}.$$

That is, we say that $p_{N_k(t_\ell)}(\alpha_k|\mathcal{B}_{t_{\ell-1}})$ is the conditional probability of signal energy being present in the kth DFT bin.

Now, we can represent the above *a priori* probability as the *rate* of the DTPP $N_k(t_\ell)$ as follows:

$$\lambda_k(t_\ell) = P\{N_k(t_\ell) = 1|\mathcal{B}_{t_{\ell-1}}\}\ k = 0,\ldots,L\ . \tag{2.10}$$

At this point, it must be emphasized that the rates $\lambda_k(t_\ell)$ do not correspond directly with the $\pi^{\alpha_1 \cdots \alpha_P}$ given above unless all sinusoids are assigned to separate DFT bins. Therefore, we may not completely describe any *a priori* probability distribution due to the resolution of the spectral analysis tools employed. Since we desire to estimate the probability structure via these marginal probabilities or rates, we conclude that we must avail ourselves of an estimator for $\lambda_k(t_\ell)$.

To do this, we require a probabilistic model to characterize this rate and a model to describe its evolution in time. A physically meaningful and, at the same time, mathematically tractable, model for $\lambda_k(t_\ell)$ is to represent it as a finite-state Markov chain. Such a model may be contrasted with a continuous model for $\lambda_k(t_\ell)$ as follows:

1. The Markov structure permits the evolution of $\lambda_k(t_\ell)$ to be treated probabilistically via the state transition matrix. This representation may be contrasted with a stochastic differential or difference equation for $\lambda_k(t_\ell)$ which may be difficult to treat analytically. The introduction of a Markov model permits the application of an exact MMSE estimator for $\lambda_k(t_\ell)$ with a recursive estimator.

2. The finite-state model permits limits on the range of $\lambda_k(t_\ell)$ to be imposed, and the rate may be restricted to the expected domain of the parameter space. Such a limitation may, for example, be chosen to reduce or eliminate the probability of runaway, which is a possibility in the decision-directed estimation context.

Under the Markov structure, we may represent $\lambda_k(t_\ell)$ as a finite-state vector Markov chain with states $\rho_1(t_\ell), \ldots, \rho_r(t_\ell)$ which can be expressed in vector form for each bin k as

$$\boldsymbol{\rho}_k(t_\ell) = \begin{bmatrix} \rho_{k,1}(t_\ell) \\ \rho_{k,2}(t_\ell) \\ \cdot \\ \cdot \\ \cdot \\ \rho_{k,r}(t_\ell) \end{bmatrix}$$

Define a Markov state vector for each bin as $\mathbf{x}_k(t_\ell) = [x_1(t_\ell), \ldots x_r(t_\ell)]^T$ where

$$x_{k,i}(t_\ell) = \begin{cases} 1 & \text{if } \lambda_k(t_\ell) = \rho_{k,i}(t_\ell) \\ 0 & \text{otherwise} \end{cases}, i = 1, \ldots, r. \tag{2.11}$$

Thus, we can represent $\lambda_k(t_\ell)$ in inner product form as

$$\lambda_k(t_\ell) = \boldsymbol{\rho}^T(t_\ell)\mathbf{x}_k(t_\ell). \tag{2.12}$$

The state, therefore, characterizes the probability distribution of the process $N_k(t_\ell)$. In other words, knowledge of $\mathbf{x}_k(t_\ell)$ specifies the probability mass function $p_{N_k(t_\ell)}(\alpha_k)$.

The evolution of the state may be characterized by means of a state probability transition matrix

$$q_{ij}(t_\ell) = P\{x_{k,j}(t_{\ell+1}) = 1 | x_{k,i}(t_\ell) = 1\} \quad i, j = 1, \ldots, r \tag{2.13}$$

Let $\mathbf{Q}(t_\ell) = [q_{ij}(t_\ell)]$ denote the state transition matrix. Then the state evolves as

$$\mathbf{x}_k(t_{\ell+1}) = \mathbf{Q}^T(t_\ell)\mathbf{x}_k(t_\ell) + \mathbf{u}_k(t_\ell) \tag{2.14}$$

where $\mathbf{u}_k(t_\ell) = \mathbf{x}_k(t_{\ell+1}) - \mathbf{Q}^T(t_\ell)\mathbf{x}_k(t_\ell)$. Define the family of σ-fields

$$\mathcal{B}_{k,t_\ell} = \sigma\{N_k(s), s \le t_\ell,\ \mathbf{x}_k(s), s \le t_{\ell+1}\};$$

we observe that $\mathbf{u}_k(t_\ell) \in \mathcal{B}_{k,t_\ell}$ and $E(\mathbf{u}_k(t_\ell)|\mathcal{B}_{k,t_{\ell-1}}) = 0$. Consequently, $\{\mathbf{u}_k\}$ is a martingale difference (MD) process with respect to the family of σ-fields $\{\mathcal{B}_k\}$ (see Appendix A). Notationally, we say $\{\mathbf{u}_k\}$ is a $\{\mathcal{B}_k\}$-MD.

Since $\lambda_k \in [0, 1]$ we must have the elements of each $\boldsymbol{\rho}_k$ vector above, bounded by unity, i.e.,

$$0 \le \rho_{k,i} \le 1, \quad k = 0, \ldots, L, \ j = 1, \ldots, r.$$

2.2.2 Estimation Procedure

A useful characterization of the process $\{N_k\}$ is to obtain its Doob decomposition with respect to $\{\mathcal{B}_k\}$. Recall (see Appendix A) that the Doob decomposition of a process $\{N_k\}$ with respect to a family of σ-fields $\{\mathcal{B}_k\}$ is the representation

$$N_k(t_\ell) = \lambda_k(t_\ell) + w_k(t_\ell)$$

where $\lambda_k(t_\ell)$ is a $\{\mathcal{B}_k\}$-predictable process (i.e., $\lambda_k(t_\ell) \in \mathcal{B}_{k,t_{\ell-1}}$ for all t_ℓ) and $w_k(t_\ell)$ is a $\{\mathcal{B}_k\}$-MD sequence (i.e., $w_k(t_\ell) \in \mathcal{B}_{k,t_\ell}$ and $E(w_k(t_\ell)|\mathcal{B}_{k,t-1}) = 0$). From the above development

$$\lambda_k(t_\ell) = E(N_k(t_\ell)|\mathcal{B}_{k,t_{\ell-1}}).$$

and if we define

$$w_k(t_\ell) = N_k(t_\ell) - E(N_k(t_\ell)|\mathcal{B}_{k,t_{\ell-1}})$$

then

$$N_k(t_\ell) = \lambda_k(t_\ell) + w_k(t_\ell) = \boldsymbol{\rho}_k^T(t_\ell)\mathbf{x}_k(t_\ell) + w_k(t_\ell) \tag{2.15}$$

is the (unique) Doob decomposition of $\{N_k\}$ with respect to $\{\mathcal{B}_k\}$

Equations (2.14) and (2.15) represent a type of state-space model for the system under study. The dynamics equation, (2.14) describes the evolution of the process $\mathbf{x}_k(t_\ell)$ over time, and is analogous to a linear difference equation driven by a noise process. The observation equation (2.15) provides the relationship of the observed process $N_k(t_\ell)$ to the state, and is analogous to the signal-in-additive-noise process familiar to linear estimation problems. Although these equations are similar to their counterparts in linear system theory, they cannot be treated the same way, since the processes $\{\mathbf{u}_k\}$ and $\{w_k\}$ are **not** additive white noise processes.

If $\mathbf{x}_k(t_\ell)$ were known, the problem of predicting $N_k(t_\ell)$ at any time t_ℓ would be solved, regardless of the past history of $N_k(\cdot)$. Unfortunately, $\mathbf{x}_k(t_\ell)$ is not directly observable; only $N_k(t_\ell)$ is observed. We are thus faced with the problem of estimating $\mathbf{x}_k(t_\ell)$ in order to render an acceptable prediction of $N_k(t_\ell)$. To formulate this estimation problem more clearly, let us define the family of σ-fields $\mathcal{F}_{k,t_\ell}$ as

$$\mathcal{F}_{k,t_\ell} = \sigma\{N_k(s), s \leq t_\ell\} \tag{2.16}$$

and compute the conditional expectation of $\mathbf{x}_k(t_\ell)$ given $\mathcal{F}_{k,t_{\ell-1}}$. To do this, we draw upon two fundamental results of martingale theory, namely, the innovations theorem and the representation theorem. These theorems are stated and discussed in Appendix A. Application of these theorems results in a nonlinear estimation procedure to obtain the Doob decomposition of $\{N_k\}$ with respect to $\{\mathcal{F}_k\}$, yielding

$$N_k(t_\ell) = \hat{\lambda}_k(t_\ell|t_{\ell-1}) + \nu_k(t_\ell) = \boldsymbol{\rho}_k^T(t_\ell)\hat{\mathbf{x}}_k(t_\ell|t_{\ell-1}) + \nu_k(t_\ell), \tag{2.17}$$

where $\{\nu_k\}$ is a $\{\mathcal{F}_k\}$-MD process and $\hat{\mathbf{x}}_k(t_\ell|t_{\ell-1})$ is the conditional expectation of $\mathbf{x}_k(t_\ell)$ given $\mathcal{F}_{k,t_{\ell-1}}$.

The process $\mathbf{x}_k(t_\ell)$ modulates the rate of the discrete-time process $N_k(t_\ell)$ according to equations (2.14) and (2.15). We wish to obtain equations of evolution of the process

$$\hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell) = E\{\mathbf{x}_k(t_{\ell+1})|\mathcal{F}_{k,t_\ell}\} = E^{\mathcal{F}_{k,t_\ell}}\mathbf{x}_k(t_{\ell+1})$$

the conditional expectation of $\mathbf{x}(t_{\ell+1})$ given the σ-field $\mathcal{F}_{k,t_\ell}$ (in standard and an alternate notation). We follow the results of [29], and obtain an estimator of the form

$$\hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell) = E^{\mathcal{F}_{k,t_{\ell-1}}}\mathbf{x}_k(t_{\ell+1}) + (\boldsymbol{\mu}_k, \nu_k)_{t_\ell}(\nu_k, \nu_k)_{t_\ell}^{-1}\nu_k(t_\ell) \tag{2.18}$$

where

$$\boldsymbol{\mu}_k(t_\ell) = E^{\mathcal{F}_{k,t_\ell}}\mathbf{x}_k(t_{\ell+1}) - E^{\mathcal{F}_{k,t_{\ell-1}}}\mathbf{x}_k(t_{\ell+1}) \tag{2.19}$$

and

$$\nu_k(t_\ell) = N_k(t_\ell) - E^{\mathcal{F}_{k,t_{\ell-1}}}N_k(t_\ell) = N_k(t_\ell) - \boldsymbol{\rho}_k^T(t_\ell)E^{\mathcal{F}_{k,t_{\ell-1}}}\mathbf{x}_k(t_\ell). \tag{2.20}$$

the matrix

$$(\boldsymbol{\mu}_k, \nu_k)_{t_\ell} = E^{\mathcal{F}_{k,t_{\ell-1}}}\boldsymbol{\mu}_k(t_\ell)\nu_k(t_\ell) \tag{2.21}$$

is the conditional covariance of $\boldsymbol{\mu}_k(t_\ell)$ and $\nu_k(t_\ell)$, and the quantity

$$(\nu_k, \nu_k)_{t_\ell} = E^{\mathcal{F}_{k,t_{\ell-1}}}\nu_k^2(t_\ell) \tag{2.22}$$

is the conditional variance of $\nu_k(t_\ell)$.

The conditional covariance $(\boldsymbol{\mu}_k, \nu_k)_{t_\ell}$ is derived in Appendix A, and is the $r \times 1$ matrix

$$\begin{aligned} E^{\mathcal{F}_{k,t_{\ell-1}}}[\boldsymbol{\mu}_k(t_\ell)\nu_k(t_\ell)] &= \mathbf{Q}^T(t_\ell)\text{diag}\,\{\hat{\mathbf{x}}_k(t_\ell|t_{\ell-1})\}\boldsymbol{\rho}(t_\ell) - \mathbf{Q}^T(t_\ell)\hat{\mathbf{x}}_k(t_\ell|t_{\ell-1})\hat{\mathbf{x}}_k^T(t_\ell|t_{\ell-1})\boldsymbol{\rho}(t_\ell) \\ &= \mathbf{Q}^T(t_\ell)[\text{diag}\,\{\hat{\mathbf{x}}_k(t_\ell|t_{\ell-1})\} - \hat{\mathbf{x}}_k(t_\ell|t_{\ell-1})\hat{\mathbf{x}}_k^T(t_\ell|t_{\ell-1})]\boldsymbol{\rho}(t_\ell) \end{aligned} \tag{2.23}$$

where diag { } denotes a diagonal matrix whose diagonal elements are composed of the elements of the vector argument.

The conditional variance $(\nu, \nu)_{t_\ell}$ is also derived in Appendix A, and is the quantity

$$E^{\mathcal{F}_{k,t_{\ell-1}}}[\nu_k^2(t_\ell)] = \hat{\lambda}_k(t_\ell|t_{\ell-1}) - \hat{\lambda}_k^2(t_\ell|t_{\ell-1}). \tag{2.24}$$

with

$$\hat{\lambda}_k(t_\ell|t_{\ell-1}) = \boldsymbol{\rho}_k^T(t_\ell)\hat{\mathbf{x}}_k(t_\ell|t_{\ell-1}). \tag{2.25}$$

Thus, the estimator becomes, using (2.23) and (2.24),

$$\hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell) = \mathbf{Q}^T(t_\ell)\hat{\mathbf{x}}_k(t_\ell|t_{\ell-1}) + E^{\mathcal{F}_{k,t_{\ell-1}}}[\boldsymbol{\mu}_k(t_\ell)\nu(t_\ell)]\left[E^{\mathcal{F}_{k,t_{\ell-1}}}[\nu_k(t_\ell)^2]\right]^{-1}\nu(t_\ell). \quad (2.26)$$

The resulting rate estimates are, therefore,

$$\hat{\lambda}_k(t_{\ell+1}|t_\ell) = \boldsymbol{\rho}_k^T(t_{\ell+1})\hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell). \quad (2.27)$$

2.2.3 Covariance of Estimation Error

The recursive estimator given by (2.26) and the corresponding rate estimate given by (2.27) represent the minimum-mean square error prediction of $\mathbf{x}_k(t_{\ell+1})$ and $\lambda_k(t_{\ell+1})$ given $\mathcal{F}_{k,t_\ell}$, the past and present data. We may obtain the conditional uncertainty on these estimates by computing the conditional covariance matrix of the estimation error

$$\begin{aligned}\tilde{\mathbf{x}}_k(t_{\ell+1}|t_\ell) &= \mathbf{x}_k(t_{\ell+1}) - \hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell)\\ \tilde{\lambda}_k(t_{\ell+1}|t_\ell) &= \lambda_k(t_{\ell+1}) - \hat{\lambda}_k(t_{\ell+1}|t_\ell)\end{aligned} \quad (2.28)$$

The estimation error covariance on $\tilde{\mathbf{x}}_k(t_{\ell+1}|t_\ell)$ may be computed as

$$\begin{aligned}P_k(t_{\ell+1}|t_\ell) &= E^{\mathcal{F}_{k,t_\ell}}\tilde{\mathbf{x}}_k(t_{\ell+1}|t_\ell)\tilde{\mathbf{x}}_k^T(t_{\ell+1}|t_\ell)\\ &= E^{\mathcal{F}_{k,t_\ell}}\mathbf{x}_k(t_{\ell+1})\mathbf{x}_k^T(t_{\ell+1}) - \hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell)\hat{\mathbf{x}}_k^T(t_{\ell+1}|t_\ell)\\ &= E^{\mathcal{F}_{k,t_\ell}}\operatorname{diag}\mathbf{x}_k(t_{\ell+1}) - \hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell)\hat{\mathbf{x}}_k^T(t_{\ell+1}|t_\ell)\\ &= \operatorname{diag}\hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell) - \hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell)\hat{\mathbf{x}}_k^T(t_{\ell+1}|t_\ell)\end{aligned}$$

and, by (2.27), we have the covariance of the rate estimation error given by

$$\begin{aligned}\Pi_k(t_{\ell+1}|t_\ell) &= E^{\mathcal{F}_{k,t_\ell}}\tilde{\lambda}_k(t_{\ell+1}|t_\ell)\tilde{\lambda}_k^T(t_{\ell+1}|t_\ell)\\ &= \boldsymbol{\rho}_k^T(t_{\ell+1})\left[\operatorname{diag}\hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell) - \hat{\mathbf{x}}_k(t_{\ell+1}|t_\ell)\hat{\mathbf{x}}_k^T(t_{\ell+1}|t_\ell)\right]\boldsymbol{\rho}(t_{\ell+1})\end{aligned}$$

At this point, some comments with regard to the recursive estimator provided in Equation (2.26) may be appropriate. Note that although this estimate is recursive and possesses structure much like a Kalman filter, it is not a linear estimator since the gain matrix is dependent upon the state and, hence, upon the data. Furthermore, the covariance associated with this estimate is a conditional covariance, rather than an unconditional covariance as with the linear case. Note also that this covariance does not obey a Riccati equation, but it is true that the (state-dependent) gain of the estimator is proportional to this covariance, as is the case with the Kalman filter. Although the estimation error covariance provided above is conditional, it may properly be used in practice to assess the quality of the estimate for $\mathbf{x}_k(t_{\ell+1})$ or $\lambda_k(t_{\ell+1})$.

2.3 Decision-Directed Detection

The DTPP estimator defined in (2.26) provides an estimate of the vector $\mathbf{x}_k(t_\ell)$, thus yielding a probability vector $\hat{\mathbf{x}}_k(t_\ell|t_{\ell-1})$ with components $\hat{x}_{k,i}(t_\ell|t_{\ell-1})$ representing the conditional probability that $\lambda_k(t_\ell) = \rho_{k,i}(t_\ell)$. The conditional expectation of $\lambda_k(t_\ell)$ given the σ-field $\mathcal{F}_{k,t_{\ell-1}}$ is

$$\hat{\lambda}_k(t_\ell|t_{\ell-1}) = \boldsymbol{\rho}_k^T(t_\ell)\hat{\mathbf{x}}_k(t_\ell|t_{\ell-1}),$$

which yields the rate estimates for each bin of the DFT.

2.3.1 Detector Design

In each of the DFT bins, two quadrature values are present and designated as the real and imaginary parts of the complex number $Y_k(t_\ell)$. When using DFT data, it is common to employ detectors using either the energy present in a bin or the square-root of the energy in a bin. The former approach gives rise to so-called envelope-squared detection and the latter approach is called envelope detection. In both instances, the original time data sequences are corrupted by Gaussian noise which transforms to Gaussian noise in each frequency bin because of the well-known property that sums of Gaussian processes are Gaussian processes. Consequently, the individual bin contents also have the form of a signal plus additive white Gaussian noise as shown in Appendix B.3. For this signal and noise description Whalen [36] gives envelope-squared and envelope detectors using χ^2 and Rician noise models respectively. In actual use, for the two-hypothesis case the form of the detectors is identical for either approach, so only the envelope-squared approach is illustrated here.

At a given time t_ℓ then, for each bin in the DFT, define the two hypotheses:

$$\mathbf{H}_{0,k}(t_\ell): \; Z_k(t_\ell) = \mu_k$$

$$\mathbf{H}_{1,k}(t_\ell): \; Z_k(t_\ell) = \beta_k$$

where $Z_k(t_\ell) = |Y_k(t_\ell)|^2$ is the energy present in a given bin, μ_k is centrally χ^2-distributed noise present in bin k, and β_k is non-centrally χ^2-distributed signal plus noise as given in

[36]. Thus, $\mathbf{H}_{0,k}(t_\ell)$ represents the noise-only case and $\mathbf{H}_{1,k}(t_\ell)$ represents the case of signal plus noise.

The likelihood ratio is expressed as a function of the *a priori* probability of each hypothesis being correct as well as the hypothesis-conditional probabilities. The decisions for the true Bayes case are then are given by the likelihood ratio test (LRT) as follows:

$$D_k[Z_k(t_\ell)] = \begin{cases} 1 & \text{if } \lambda_k(t_\ell)p(Z_k(t_\ell)|\mathbf{H}_{1,k}(t_\ell)) \geq (1-\lambda_k(t_\ell))p(Z_k(t_\ell)|\mathbf{H}_{0,k}(t_\ell)) \\ 0 & \text{otherwise} \end{cases} . \quad (2.29)$$

where the conditional probabilities $p(Z_k(t_\ell)|\mathbf{H}_{0,k}(t_\ell))andp(Z_k(t_\ell)|\mathbf{H}_{1,k}(t_\ell))$ are represented by central and non-central χ^2 distributions respectively. The empirical Bayes philosophy is to use estimates of the prior probabilities in place of the actual probabilities. If the rate estimates, $\hat{\lambda}_k(t_\ell)$ are used and estimates for the parameters present in the conditional probabilities are employed, we say that this is a *generalized* likelihood ratio test (GLRT):

$$\hat{D}_k[Z_k(t_\ell)] = \begin{cases} 1 & \text{if } \hat{\lambda}_k(t_\ell)\hat{p}(Z_k(t_\ell)|\mathbf{H}_{1,k}(t_\ell)) \geq (1-\hat{\lambda}_k(t_\ell))\hat{p}(Z_k(t_\ell)|\mathbf{H}_{0,k}(t_\ell)) \\ 0 & \text{otherwise} \end{cases} . \quad (2.30)$$

The complete derivation of the likelihood ratio is presented in Appendix B, but the result is that we can construct the log-likelihood function as a threshold τ_k for each bin as

$$\tau_k = \log\hat{\lambda}_k - \frac{\hat{\gamma}_k}{2} + \log(I_0((q_k\hat{\gamma}_k)^{1/2}) - \log(1-\hat{\lambda}_k) \quad (2.31)$$

where I_0 is the modified Bessel function of the first kind, order 0, and γ_k is the noncentral parameter defined as $\gamma_k = 2\zeta_k^2$ where ζ_k is the signal envelope amplitude for bin k. The quantity q_k is the χ^2 distributed observation defined as:

$$q_k = \frac{f_{1,k}^2}{\hat{\sigma}^2} + \frac{f_{2,k}^2}{\hat{\sigma}^2}.$$

for $f_{1,k}$ = Real $Y_k(t_\ell)$ and $f_{2,k}$ = Imaginary $Y_k(t_\ell)$. Thus,

$$n_k(t_\ell) = \begin{cases} 1 & \text{if } \tau_k \geq 1 \\ 0 & \text{otherwise} \end{cases} .$$

Typically, in order to more efficiently compute the detection threshold, we would employ small- and large-argument approximations for the modified Bessel function and use these values in the log-likelihood ratio test. Since even the approximations desired are in exponential form, the τ_k values arrived at are usually just summations. In a system

realization then, there is a bank of $L = \frac{M}{2} + 1$ scalar detectors of the form described above which generate a vector of binary detector decisions which we may express as

$$\mathbf{n}(t_\ell) = \begin{bmatrix} n_0(t_\ell) \\ n_1(t_\ell) \\ \cdot \\ \cdot \\ \cdot \\ n_L(t_\ell) \end{bmatrix}$$

This is illustrated in Figure 2.4.

2.3.2 Bias Correction

Unfortunately, it is not generally true that the $\hat{\lambda}_k(t_\ell)$ is an unbiased estimate, and we must investigate the effects of this bias on the performance of the detector and, if necessary, explore methods of eliminating or reducing this bias.

For any partitioning $\mathcal{T}$, of the decision space, we may express $\lambda^{\alpha_k}(t_\ell)$ in terms of $\pi^{\alpha_k}(t_\ell)$ as

$$\begin{aligned} \lambda^{\alpha_k}(t_\ell) = \; & \pi^{\alpha_k}(t_\ell) \int_{T^{\alpha_k}} \left[f(Z_k(t_\ell)|\mathbf{H}_{\alpha_k,k}(t_\ell)) - f(Z_k(t_\ell)|\overline{\mathbf{H}}_{\alpha_k,k}(t_\ell)) \right] dZ_k(t_\ell) \\ & + \int_{T^{\alpha_k}} f(Z_k(t_\ell)|\overline{\mathbf{H}}_{\alpha_k,k}(t_\ell)) dZ_k(t_\ell) \end{aligned}$$

and, solving for $\pi^{\alpha_k}(t_\ell)$ yields

$$\pi^{\alpha_k}(t_\ell) = \frac{\lambda^{\alpha_k}(t_\ell) - \int_{T^{\alpha_k}} f(Z_k(t_\ell)|\overline{\mathbf{H}}^{\alpha_k,k}(t_\ell)) dZ_k(t_\ell)}{\int_{T^{\alpha_k}} \left[f(Z_k(t_\ell)|\mathbf{H}^{\alpha_k,k}(t_\ell)) - f(Z_k(t_\ell)|\overline{\mathbf{H}}^{\alpha_k,k}(t_\ell)) \right] dZ_k(t_\ell)} = a(T^{\alpha_k})\lambda^{\alpha_k}(t_\ell) + b(T^{\alpha_k}) \tag{2.32}$$

where $a(\cdot)$ and $b(\cdot)$ are defined in an obvious way. Thus, the true rate π is expressed as a linear function of the detected rate, λ. In general, $a(\cdot) \neq 1$ and $b(\cdot) \neq 0$. However, for any given decision region T^{α_k}, the correction terms may be computed and applied. If we estimate $\lambda^{\alpha_k}(t_\ell)$ using the above scheme, we may then express the estimate of $\pi^{\alpha_k}(t_\ell)$ as

$$\hat{\pi}^{\alpha_k}(t_\ell) = a(T^{\alpha_k})\hat{\lambda}^{\alpha_k}(t_\ell|t_{\ell-1}) + b(T^{\alpha_k}).$$

This structure holds for all values of T^{α_k} and, in particular, holds when the partition regions T^{α_k} are specified by the previous best estimate of the prior, namely, $\hat{\pi}(t_{\ell-1})$.

There are a number of issues to be considered concerning the removal of the bias. First, it is evident from the structure of (2.32) that $a(\cdot) \geq 1$, and therefore the variance on the estimation error for π^{α_k} will be greater than the variance on the estimation error for λ^{α_k}, thus bias may be removed only at the expense of increased uncertainty in the estimate. Second, the integrations indicated in (2.32) are extremely complex, since the integrations are taken over m-dimensional space. For the Gaussian case, these integrals cannot be evaluated in closed form, and the computational burden to numerically evaluate them is considerable. Consequently, for the present analysis, we simply neglect the bias and use the simple estimator defined by (2.25). In Chapter 3, Monte Carlo results are provided to provide partial justification for this simplifying procedure.

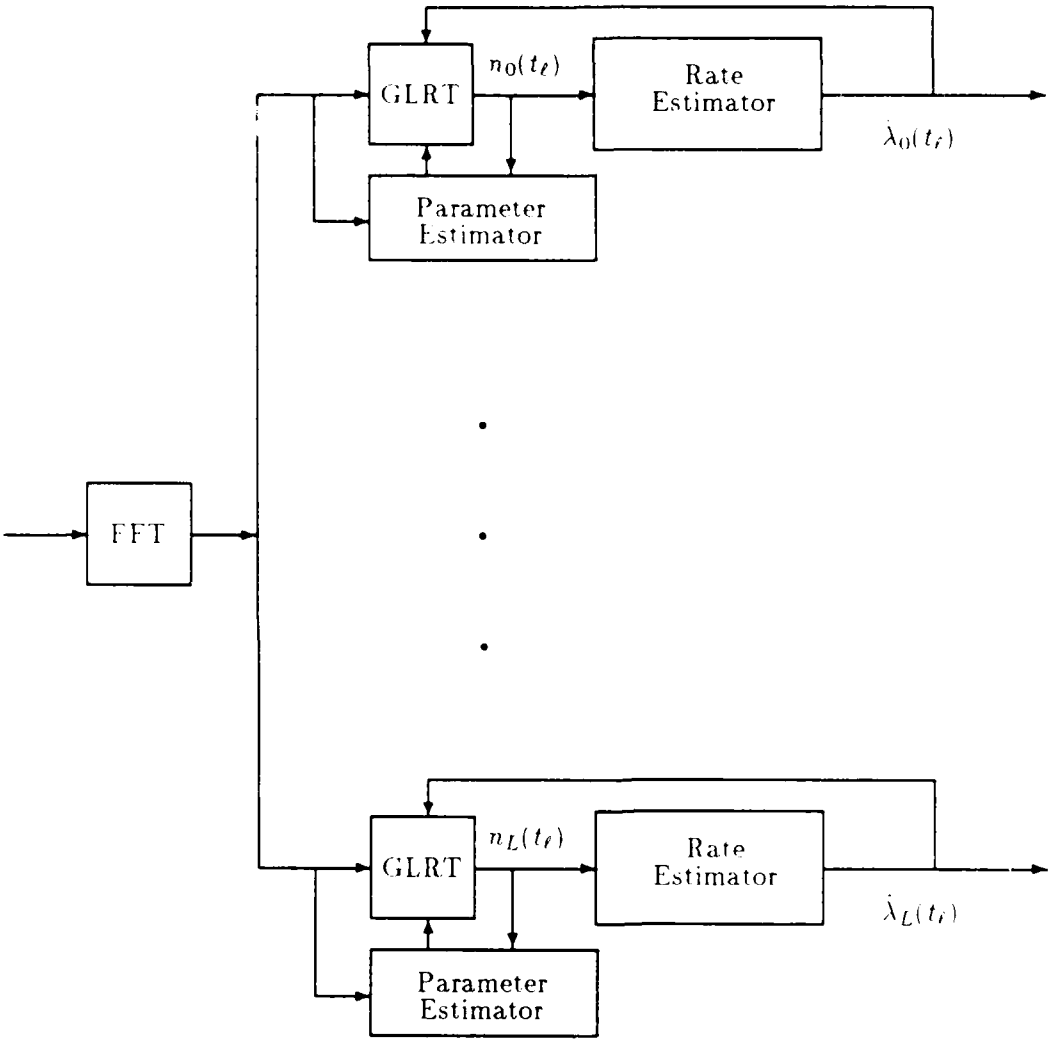


Figure 2.4: Decision-Directed Spectral Probability Estimation Block Diagram

Chapter 3

Performance Evaluation by Monte Carlo Analysis

A key analysis issue is to assess the performance of the proposed algorithm. The interaction between detection and estimation, however, makes performance analysis of this decision-directed procedure extremely complex using classical procedures. The difficulty is due primarily to the dependencies present in the adaptive detector that are introduced by the two-way coupling between detection and estimation, which are virtually impossible to treat since the multivariate distributions are not available in analytic form. An alternative to a classical performance analysis is to conduct Monte Carlo analyses, and to evaluate the performance of this algorithm on the basis of first and second sample moments of the trial results. To this end, we present simulation results to demonstrate the operation of the algorithms presented in Chapter 2. In addition to using estimates of the prior to specify the decision rule, it is also necessary to estimate certain unknown signal parameters such as signal strength, which results in a further generalization of the likelihood ratio test.

3.1 Signal Modeling Assumptions

The specification of the parameters of the controlled experiments to be conducted under this study require that a signal model be supplied to generate the observations. This signal model consists of a time history of the marginal probabilities for each frequency bin. Furthermore, it is also necessary to specify the structure of the Markov chain model used to estimate the prior, and to establish procedures for decision-directed estimation of the

signal strength and the noise variance.

3.1.1 Observations Model

For all simulations done in order to arrive at Monte Carlo performance estimates, the size of the analysis blocks is chosen to be 32 samples. This means that the size of the DFT (FFT) is $M = 32$ bins wide, and that the frequency detection is performed on the bins indexed 0 through 16. Any frequency data in bins 17 through 31 is redundant. Larger block sizes are possible to implement since the computational burden grows linearly with relation to the transform size. It is easier to run Monte Carlo simulations, however, when the size is relatively small. Energy detection is then done for each bin via the procedures outlined in Chapter 2 and the DTPP describing the detector decisions drives the estimator for the *a priori* probability.

In general, we assume the signal model presented in Chapter 2, namely: Let $y(t)$ be a received signal composed of possibly harmonically related sinusoids contaminated with additive white Gaussian noise, expressed as

$$y(t) = \sum_{j=1}^{P} a_j(t) A_j(t) \cos(\omega_j t + \phi_j) + \nu(t), \tag{3.1}$$

where ω_j is the frequency of signal j, ϕ_j is the phase angle for signal j, uniformly distributed from 0 to 2π, $A_j(t)$ is the time-varying amplitude of signal j, and $\nu(t)$ is additive Gaussian white noise. The $a_j(t)$ are discrete-time point processes (DTPP's) describing the signal presence at time t.

As previously expressed in Chapter 2, it is to be expected that, since the resolution of the FFT is not arbitrarily fine, the sinusoids given in Equation (3.1) will not necessarily be separated into distinct bins. Thus, the model which more explicitly fits the analysis scenario to which we have committed is given as follows:

$$y(t_\ell) = \sum_{k=0}^{L} n_k(t_\ell) \tilde{A}_k(t_\ell) \cos(\omega_k t_\ell + \phi_k) + \nu(t_\ell) \text{ for } L = \frac{M}{2}. \tag{3.2}$$

where k is the bin index of the FFT, ω_k is the center frequency of bin k, ϕ_k is the phase angle associated with bin k, $\tilde{A}_k(t_\ell)$ is the time-varying amplitude of the signal for bin k, and $\nu(t_\ell)$ is overall additive white Gaussian noise. Since this model differs from that given

in Equation (3.1), a different DTPP is defined in order to properly distinguish between the two models. Thus, the new DTPP modeling the signal presence is defined as $n_k(t_\ell)$. As before,

$$n_k(t_\ell) = \begin{cases} 1 & \text{if signal present in bin } k \text{ at time } t_\ell \\ 0 & \text{otherwise} \end{cases}$$

Also, let us define the detection DTPP, $N_k(t_\ell)$, as

$$N_k(t_\ell) = \begin{cases} 1 & \text{if signal detected in bin } k \text{ at time } t_\ell \\ 0 & \text{otherwise} \end{cases}$$

In our processing model, we perform the analysis of the data in terms of blocks; hence, the time indices for the various quantities are, of necessity, discrete and quantized. Therefore, in order to avoid confusion with simple discrete time (the sampling instances), the subscript ℓ is added to the time variable to indicate the particular analysis block which is being processed.

We also need to define the σ-fields which influence the detector decisions. For a fundamental description of σ-fields, see the explanation of martingale theory. Suppose that $\mathcal{B}_{t_\ell}$ is the σ-field generated by all past decisions and all past detector rates. We can express this as

$$\mathcal{B}_{t_\ell} = \sigma(\mathbf{N}(s), s \leq t_\ell \; \tilde{\boldsymbol{\lambda}}(s), \; s \leq t_{\ell+1}), \tag{3.3}$$

where $\tilde{\boldsymbol{\lambda}}(s)$ is the true detection rate at time s, and $\mathbf{N}(s)$ is the vector of detector decisions at time s.

In order to fix ideas concerning the probability structures in question, let us define the probability structure governing the DTPP is Equation (3.1) in the following terms:

$$\pi_j(t_\ell) = P(\text{signal is present at frequency } \omega_j \text{ at time } t_\ell). \tag{3.4}$$

The probability structure for Equation (3.2) will be defined to be:

$$\lambda_k(t_\ell) = P(\text{signal is present in FFT bin } k \text{ at time } t_\ell). \tag{3.5}$$

And define the detection probability structure to be:

$$\tilde{\lambda}_k(t_\ell) = P(\text{signal is detected in FFT bin } k \text{ at time } t_\ell | \mathcal{B}_{t_\ell}). \tag{3.6}$$

The differences between these three probabilities are obvious, and yet the distinctions can sometimes be lost in the mass of notation which often must be employed in attempting to clarify the issues involved. In more succinct notation, and shifting to the point process aspect of the problem, we can define the above three quantities in terms of the rates of the point processes given in Equations (3.1) and (3.2) and the quantity $N_k(t_\ell)$, the detections. Let

$$\pi_j = P(a_j(t_\ell) = 1),$$

$$\lambda_k(t_\ell) = P(n_k(t_\ell) = 1), \text{ and} \tag{3.7}$$

$$\tilde{\lambda}_k(t_\ell) = P(N_k(t_\ell) = 1|\mathcal{B}_{t_\ell}).$$

It is important to note that the σ-field $\mathcal{B}_{t_\ell}$ is not observed since, even though we know the detection process, we will not, in general, ever know $\tilde{\lambda}_k(t_\ell)$, the true rate of detections. Thus, we let $\mathcal{F}_{t_\ell}$ be the σ-field generated by the detections as

$$\mathcal{F}_{t_\ell} = \sigma(\mathbf{N}(s), \ \ s \leq t_\ell), \tag{3.8}$$

where $\mathbf{N}(s)$ is the vector of detector decisions at time s. Since we do not have the true detection rates $\tilde{\lambda}_k(t_\ell)$, we let the estimated rate be $\hat{\lambda}_k(t_\ell)$. We then express the estimate $\hat{\lambda}(t_\ell)$ as

$$\hat{\lambda}_k(t_\ell) = P(N_k(t_\ell) = 1|\mathcal{F}_{t_\ell}). \tag{3.9}$$

This rate is the estimated *a priori* probability for use in the generalized likelihood ratio test. Thus, the empirical Bayes description of the processing algorithms [30]. The results in Equation 3.9 are based upon the foundation of the Doob decomposition as presented in Section 2.2.2 and Appendix A.

In the simulations then, we consider the estimation of all parameters included in Equation (3.2) except ω_k - this parameter is implicitly estimated by the detection of the signal in the kth bin. In the process of estimating the various parameters, it will be helpful to define another set of parameters used exclusively in the estimation procedure, but which relate back to the parameters in the model given in Equation (3.2). Define these as follows:

- $\hat{A}_k(t_\ell)$ - an estimate of $\tilde{A}_k(t_\ell)$.
- $\hat{\sigma}_k$ - an estimate of the variance of the noise $\nu(t_\ell)$.

- $\hat{\lambda}_k(t_\ell)$ - an estimate of the detection probability $\tilde{\lambda}_k(t_\ell)$.

In the Monte Carlo simulations, a time history for the probability structure on $\lambda_k(t_\ell)$ is defined for $k = 0,\ldots,16$ and $i = 0,\ldots,60$. For each Monte Carlo run we process 960 time points. The probability structure is generally the form of a signal swept from low frequency to higher frequency with some harmonic content. As information needs to be gathered for several signal-to-noise ratios, the simulations are set to run 10 Monte Carlo trials at each SNR as the SNR is varied from about 12 dB to -12 dB in 16 steps. At each SNR, the probability of error is computed and output. These data points constitute the receiver-operating characteristics and are plotted to show the detector performance. Also, the estimated rates for the various DTPP's are plotted as a probability surface. These plots demonstrate the algorithm's ability to track the defined *a priori* probability structure.

It should be noted that the SNR is measured with respect to the time-domain signal and therefore does not reflect any bandwidth measurement. Thus, the actual SNR in a given bin is higher than the SNR in the time-domain by a factor proportional to the number of points in the analysis window. We can explain this by the fact that the noise power is evenly distributed into the various bins while the signals are positioned within only one or two bins. Thus, comparisons with techniques reporting performance with respect to SNR/Hz are not directly possible for this research.

3.1.2 Markov Chain Model

In our estimation procedure for the rates of the various DTPP's, we invoke a Markov model to describe the time evolution of the rates as described in Chapter 2. Here we also assume that all FFT bins will be governed by the identical Markov transition matrices $\mathbf{Q}$ and states described as $\boldsymbol{\rho}$. This simplifying assumption is not very restrictive since sufficient dimension is allowed to estimate the state of the Markov chain for a given bin. It is well to note that the state estimators for the state vectors $\mathbf{x}_k$ are all done independently so that there is no coupling among rate estimators. This is partially justified since in Appendix B.3 we demonstrate that the noise is uncorrelated between FFT bins.

We now define the necessary quantities required to use Equations 2.12 and 2.14. For

the simulations we will set $r = 7$, and define the vector $\boldsymbol{\rho}$ and the matrix $\mathbf{Q}$ as

$$\begin{aligned}\boldsymbol{\rho}^T &= [\rho_1\ \rho_2\ \rho_3\ \rho_4\ \rho_5\ \rho_6\ \rho_7] \\ &= [.05\ .20\ .40\ .60\ .70\ .80\ .90],\end{aligned}$$

and

$$\mathbf{Q}^T = \begin{bmatrix} .95 & .02 & .01 & .01 & .01 & .00 & .00 \\ .02 & .93 & .02 & .01 & .01 & .01 & .00 \\ .01 & .02 & .92 & .02 & .01 & .01 & .01 \\ .01 & .01 & .02 & .92 & .02 & .01 & .01 \\ .01 & .01 & .01 & .02 & .92 & .02 & .01 \\ .00 & .01 & .01 & .01 & .02 & .93 & .02 \\ .00 & .00 & .01 & .01 & .01 & .02 & .95 \end{bmatrix}.$$

The values ρ_j, $j = 1, \ldots, 7$, are the states of the vector Markov chain, and represent the states to which the rate $\hat{\lambda}_k(t_\ell)$ may transit as time evolves. Since the states of the Markov chain represent probabilities, the only real constraint on the states is that $0 \leq \rho_j \leq 1$. Since, in general, the estimator employed to obtain $\hat{\mathbf{x}}_k$ will have non-zero values for all elements, the relatively small dimension of the states is not a severe restriction since the rate estimate will actually lie on the convex closure of the states defined in the $\boldsymbol{\rho}_k$ vector. In other words, $\hat{\lambda}_k(t_\ell)$ is a convex linear combination of the states ρ_j.

An element q_{ij} of $\mathbf{Q}^T$ represents the probability of transiting to state j of the Markov chain at time $t_{\ell+1}$ given that the state was i at time t_ℓ i.e,

$$q_{ij} = P(x_j(t_{\ell+1}) | x_i(t_\ell)).$$

As an illustration of this idea, note that the strong diagonal structure of $\mathbf{Q}^T$ indicates that given the Markov chain is in state j at time t_ℓ, it will likely remain there at $t_{\ell+1}$. Therefore, the evolution of the Markov chain is described by the equation as given in Equation 2.14, namely

$$\mathbf{x}_k(t_{\ell+1}) = \mathbf{Q}^T \mathbf{x}_k(t_\ell) + \mathbf{u}_k(t_\ell).$$

This can also be modified to increase the off-diagonal elements in order to make the system more responsive to rapid transitions if the operational scenario warrants this assumption.

3.1.3 Estimation of Signal Parameters

The signal strength for the detector structure we employ is also known as the envelope magnitude. We may not assume that the measure of signal strength in bin k, $\tilde{A}_k(t_\ell)$,

is known *a priori*, so it is important to investigate how one might generate an estimate $\hat{A}_k(t_\ell)$ of $\tilde{A}_k(t_\ell)$. Probably the most straight-forward method, and one that has previously been used with success in [18,30], is to compute a decision-directed estimate $\hat{A}_k(t_\ell)$, as the empirical average of those samples $Y_k(t_\ell)$ for which a detection in bin k occurred (i.e., when S_k occurs). This estimator assumes the general form

$$\hat{A}_k(t_\ell) = \hat{A}_k(t_{\ell-1}) + \frac{N_k(t_\ell)}{\sum_{s=1}^{t_\ell} \mu^{t-s} N_k(s)} [|Y_k(t_\ell)| - \hat{A}_k(t_{\ell-1})], \tag{3.10}$$

where the μ is a constant such that $0 < \mu \leq 1$. This quantity represents a "forgetting factor," which permits earlier estimates to be discounted in favor of more recent data. Using such a model, smooth changes in $\tilde{A}_k(t_\ell)$ may be tracked.

In order to estimate the variance of the Gaussian noise, $\nu(t_\ell)$, we employ a slightly different decision-directed approach. Since the variance we wish to estimate is contained in the raw FFT bin data, the bins useful for estimating the noise variance are those in which no signal is detected. Also, for our signal-plus-noise model, the noise variance should be equal in all bins since we assume that the noise has a white spectrum. So we compute an estimate noise variance for each eligible bin, then average across all eligible bins to arrive at the global noise variance estimate $\hat{\sigma}^2(t_\ell)$. The estimator for each bin is similar to Equation (3.10) and is given as:

$$\hat{\sigma}_k^2(t_\ell) = \hat{\sigma}_k^2(t_{\ell-1}) + \frac{[1 - N_k(t_\ell)]}{\sum_{s=1}^{t_\ell} \mu^{t_\ell - s}[1 - N_k(s)]} [|Y_k(t_\ell)|^2 - \hat{\sigma}_k^2(t_{\ell-1})] \tag{3.11}$$

Note that the recursions presented in Equations (3.10) and (3.11) must be initialized with some *a priori* estimate for $\hat{A}_k(0)$ and $\hat{\sigma}^2(0)$. Furthermore, the ratios

$$\frac{N_k(t_\ell)}{\sum_{s=1}^{t} \mu^{t_\ell - s} N_s} \quad \text{and} \quad \frac{[1 - N_k(t_\ell)]}{\sum_{s=1}^{t} \mu^{t_\ell - s}[1 - N_k(s)]}$$

must be initialized to zero at $t_\ell = 0$ to ensure that the estimates for $A_k(t_\ell)$ and $\sigma_k^2(t_\ell)$ are well-defined, and are equal to the *a priori* values until observations are obtained.

3.2 Spectral Probability Estimation

Consider the problem of estimating the probability of signal presence in a given frequency range when the received time-domain signal is obtained from either a beamformed array output, or the output of a single transducer such as a sonobuoy. Consequently, the time-domain received signal is given by Equation (3.1), and after the M-point FFT has been performed on the data for block i, we have the frequency-domain data given in vector form as $\mathbf{Y}(t_\ell)$ or

$$\mathbf{Y}(t_\ell) = \begin{bmatrix} Y_0(t_\ell) \\ Y_1(t_\ell) \\ \cdot \\ \cdot \\ \cdot \\ Y_L(t_\ell) \end{bmatrix}$$

where $L = \frac{M}{2}$. Due to the nature of the scenario, the above observations, whether in the time or frequency domain, are implicitly functions of spatial positioning, but for the present we omit any reference to the spatial dependence of the problem.

Thus, for this problem we directly apply the analysis presented in the preceding pages to arrive at the estimates for the marginal probabilities necessary for each frequency bin.

3.3 Monte Carlo Simulation Results

3.3.1 Time-varying Probability Tracking

Essentially two cases are examined in the Monte Carlo simulations. Case one, where the only the rates are estimated for use in the GLRT and case two, where the rates and the amplitudes are estimated for the GLRT. In all simulations for both cases, the same true time-varying *a priori* probability structure was used. This is shown as a three-dimensional plot in Figure 3.1. In this figure , the blocked time index runs from the back to the front of the probability surface, and the FFT bin index runs from left to right. Probability is measured perpendicularly to the plane according to the scale in the left-hand corner of the surface. Thus the figure illustrates a high probability of signal presence in the lower frequencies during the first few time blocks. The high-probability frequencies then are increased as time progresses. Ultimately, the bin with the highest probability is bin 6 at

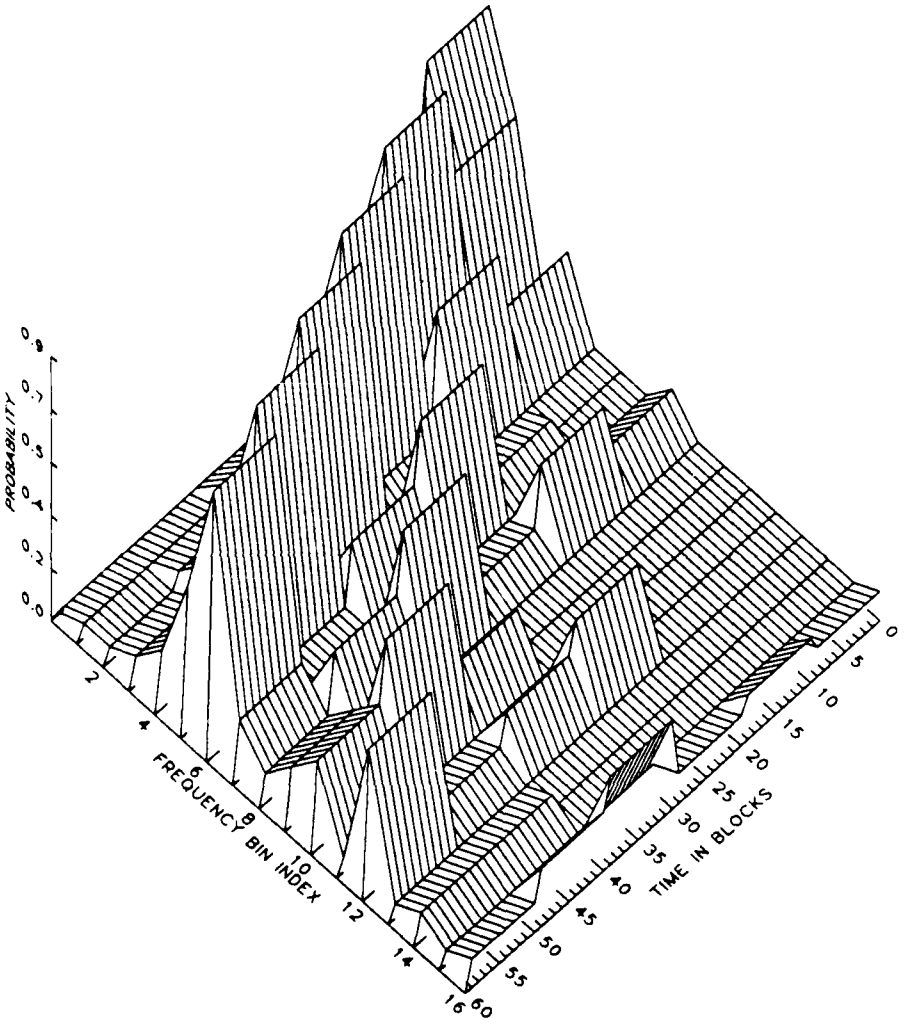


Figure 3.1: True Time-varying *A Priori* Probability Structure.

time block 60. The probability surface also illustrates harmonic structure in bins indexed as integral multiples of the dominant frequency bin. Some alternate ways of explaining the plot are to consider the representation of the data as either the collection of the marginals presented as slices as time evolves, or as a side-by-side collection of the time evolution of each marginal. All of the subsequent three-dimensional plots are constructed in the same manner with the requisite quantity plotted against time and frequency.

In Figure 3.2 we show the behavior of the rate estimator for the case where signal amplitude and noise variance are assumed known. This clearly illustrates the tracking performance of the algorithm when the probability structure is time-varying. This estimate was performed at a signal-to-noise ratio of 12 dB.

Figure 3.3 is the average of the rate estimates done for various Monte Carlo simulations and is much smoother than the single trial estimate shown in Figure 3.2. In this figure, the true prior probability structure is more obvious due to the smoothing of the average. Thus, it is evident that, in the mean, the algorithm is able to track time-varying probability structures. The confidence in the estimate can be expressed by examining the variance of the estimator as well as the mean-square error for the estimator. These quantities are plotted in Figures 3.4 and 3.5 respectively. We see from Figure 3.4 that the variance computed from the Monte Carlo procedure is never greater than 0.0487, and generally is between 0.01 to 0.02. In Figure 3.5, we see that the mean-square error is somewhat larger along the high-probability frequency track than in the low-probability regions of the spectrum. That is, the large values range from 0.5 to 0.6, but there are only about five points in the plot where these values are present. For most of the rest of the spectrum, the mean-square error is less than 0.1. The reason that the errors are large in these high-probability areas is because the estimator is adaptive. In other words, the estimates for the various probabilities are causal estimates performed in real time; no smoothing is being performed. Notice that the points at which the probabilities shift to the next bin are the points at which the mean-square error is large. After the shift, the probability estimator adapts and the mean-square error gradually decreases to a somewhat lower level. Both the variance and mean-square error results were to be expected from similar results obtained in [34].

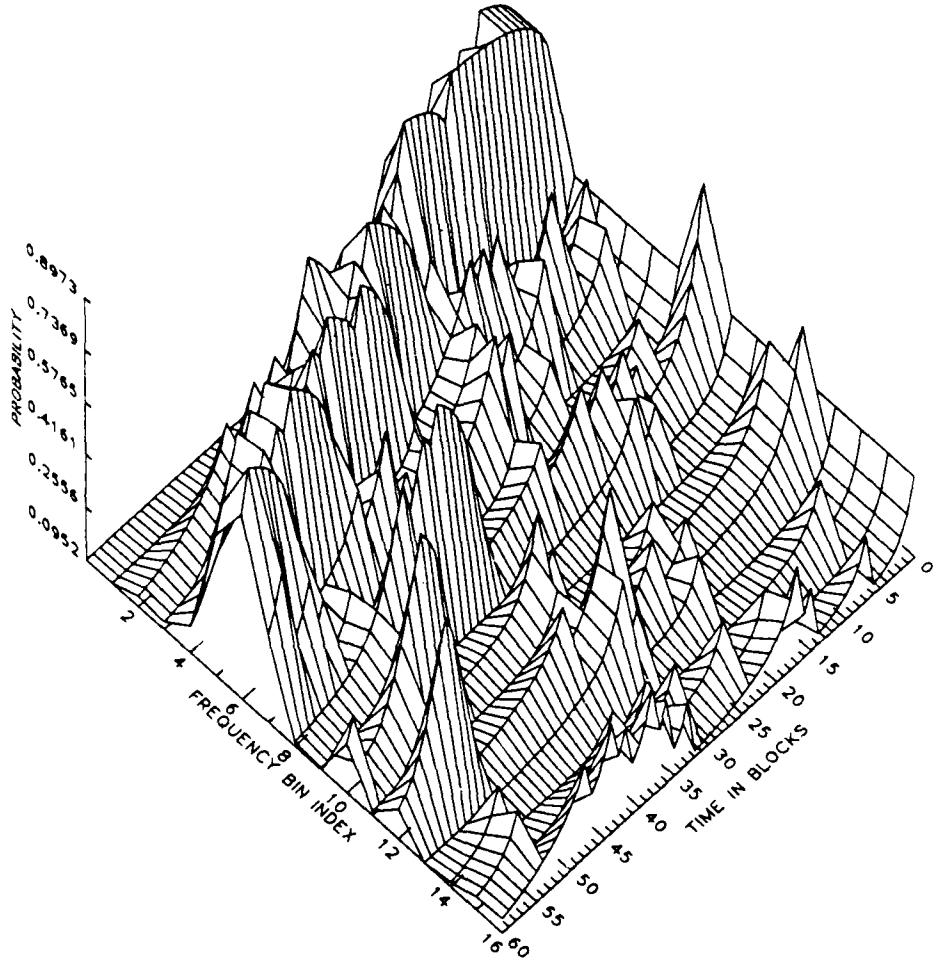


Figure 3.2: Estimated Time-Varying *A Priori* Probability Structure (SNR=12dB).

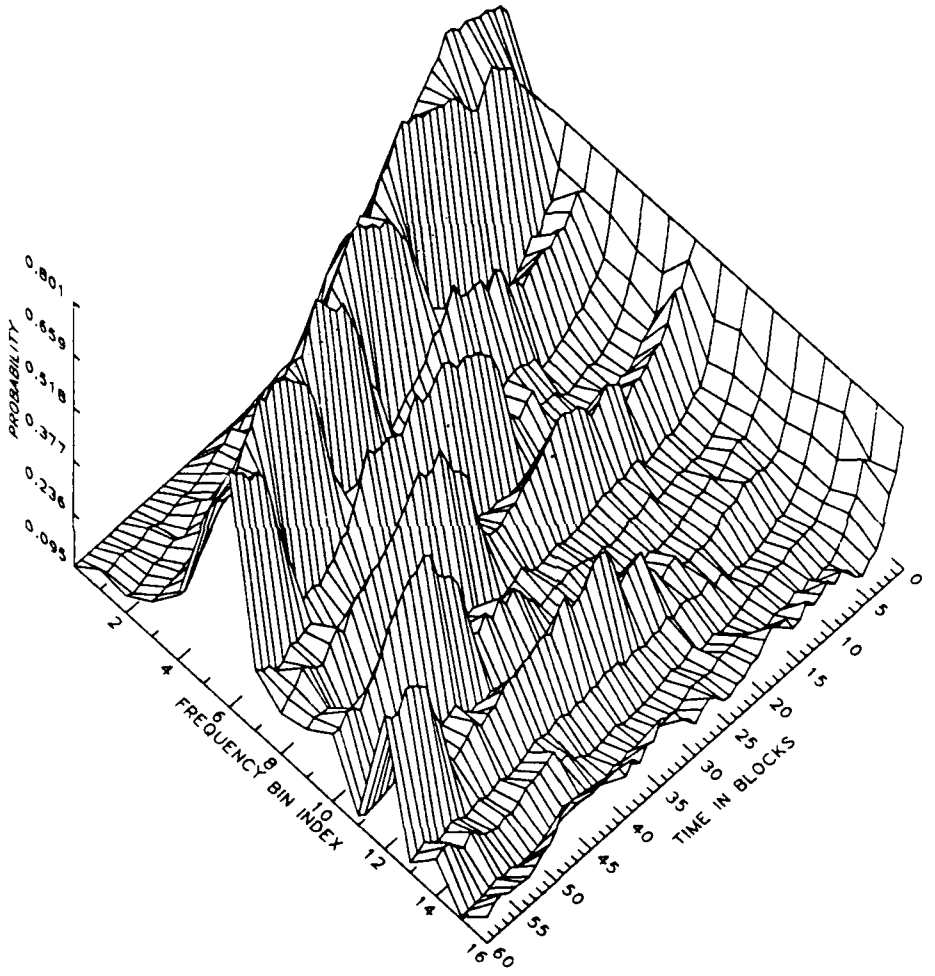


Figure 3.3: Monte Carlo Average for the Estimated *A Priori* Probability Structure (SNR=12dB).

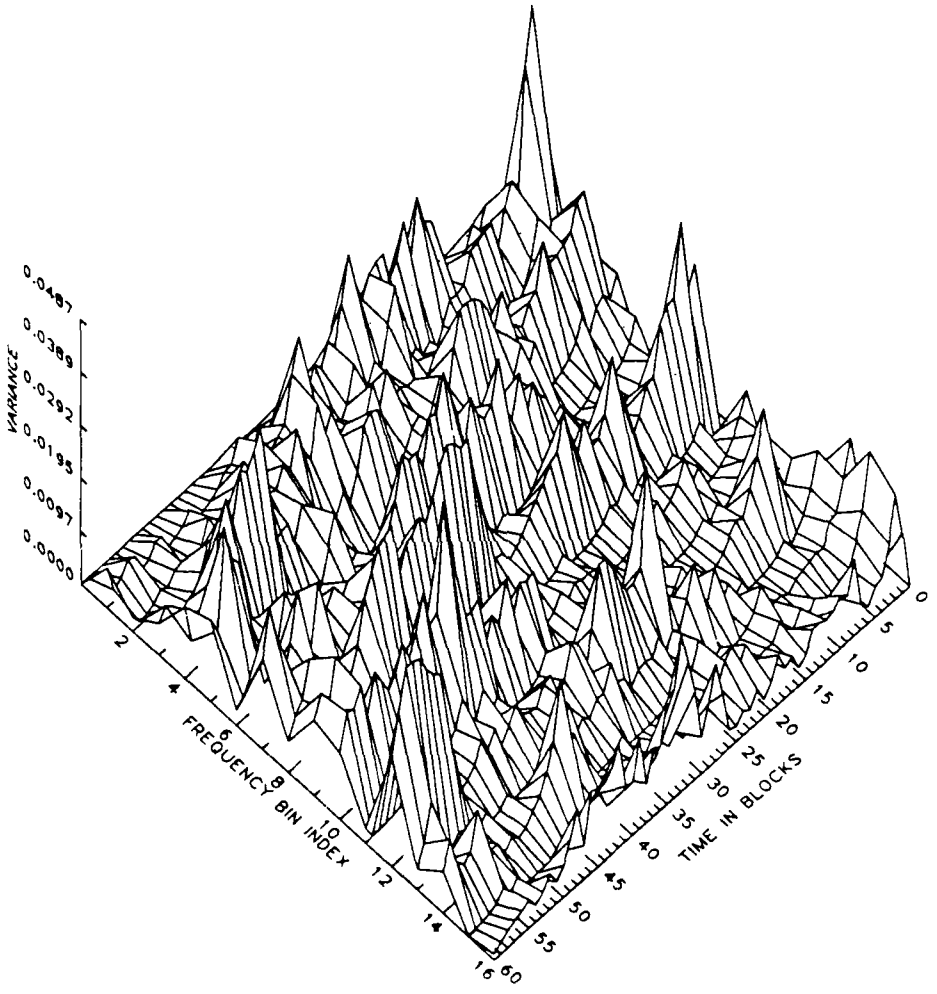


Figure 3.4: Monte Carlo Variance on the Estimated *A Priori* Probability Structure (SNR=12dB).

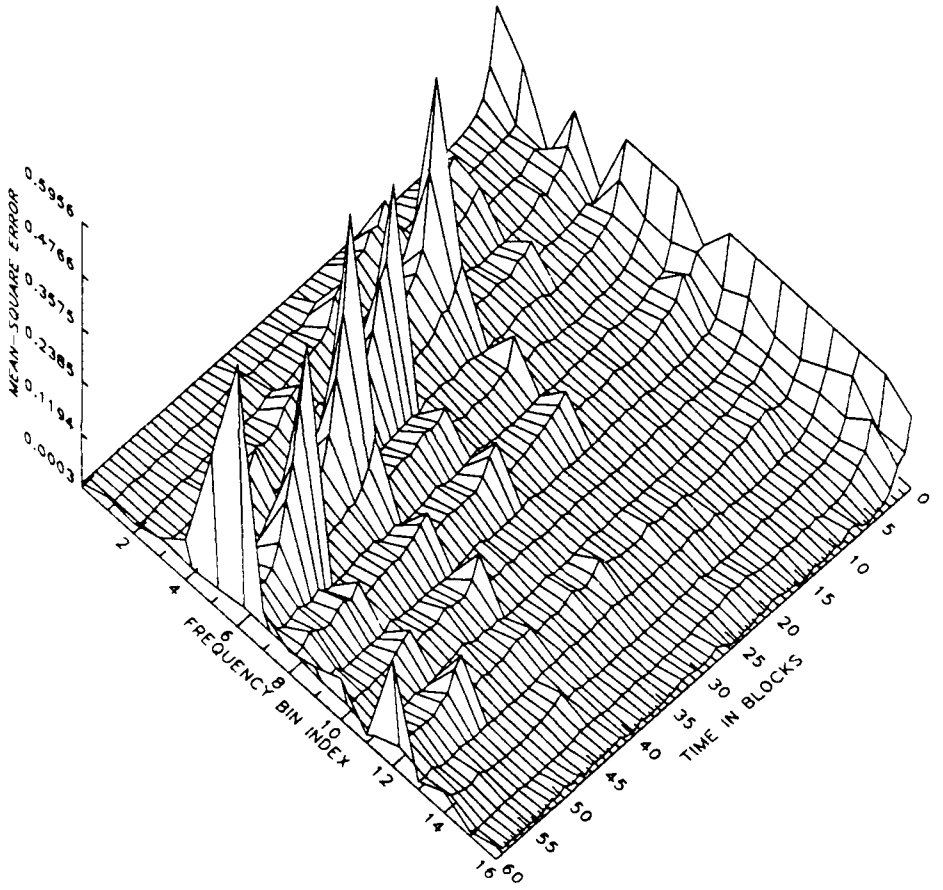


Figure 3.5. Mean-square Error for the Estimated *A Priori* Probability Structure (SNR = 12dB)

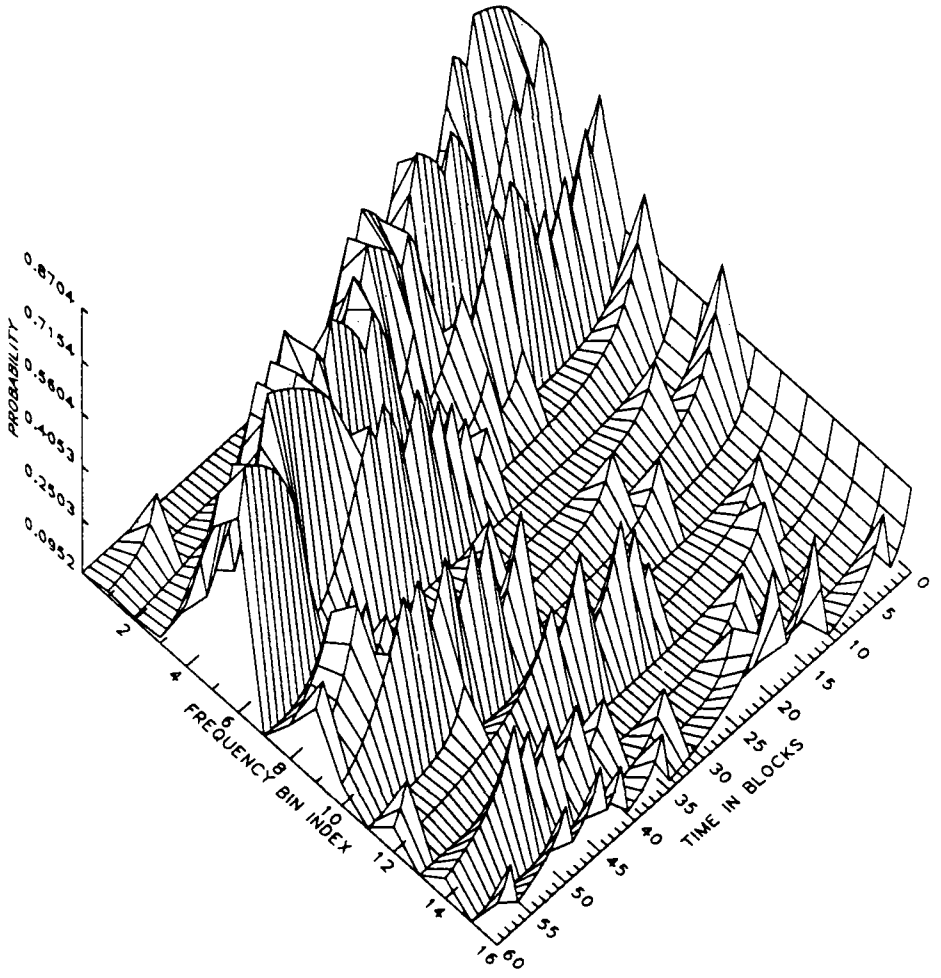


Figure 3.6: Estimated *A Priori* Probability Structure (SNR=6dB).

Next, we show results similar to Figures 3.2 - 3.5 for a signal-to-noise ratio of 6 dB. These results are still for the case where only the rate is being estimated and are presented in Figures 3.6 - 3.9. As is to be expected, the detection probability, while still quite adequate, is a little less responsive to tracking the time-varying probability in Figure 3.6; and the Monte Carlo averages in Figure 3.7 are not quite as accurate a representation of the true *a priori* probabilities as was presented for the 12 dB SNR in Figure 3.3. The variance and mean-square error presented in Figures 3.8 and 3.9 respectively are also greater than in the plots for the 12 dB simulations.

Now, we examine case two, where the rates and the bin signal amplitudes are estimated

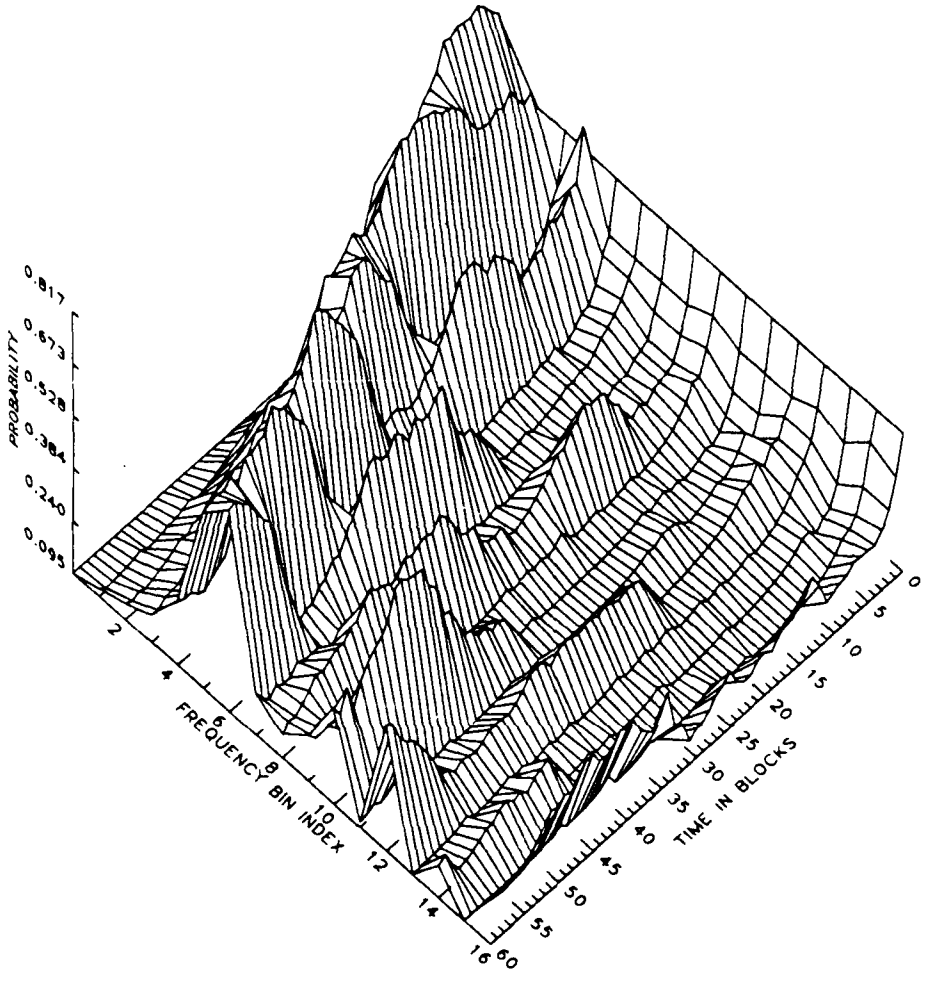


Figure 3.7: Monte Carlo Averages for the Estimated *A Priori* Probability Structure (SNR=6dB).

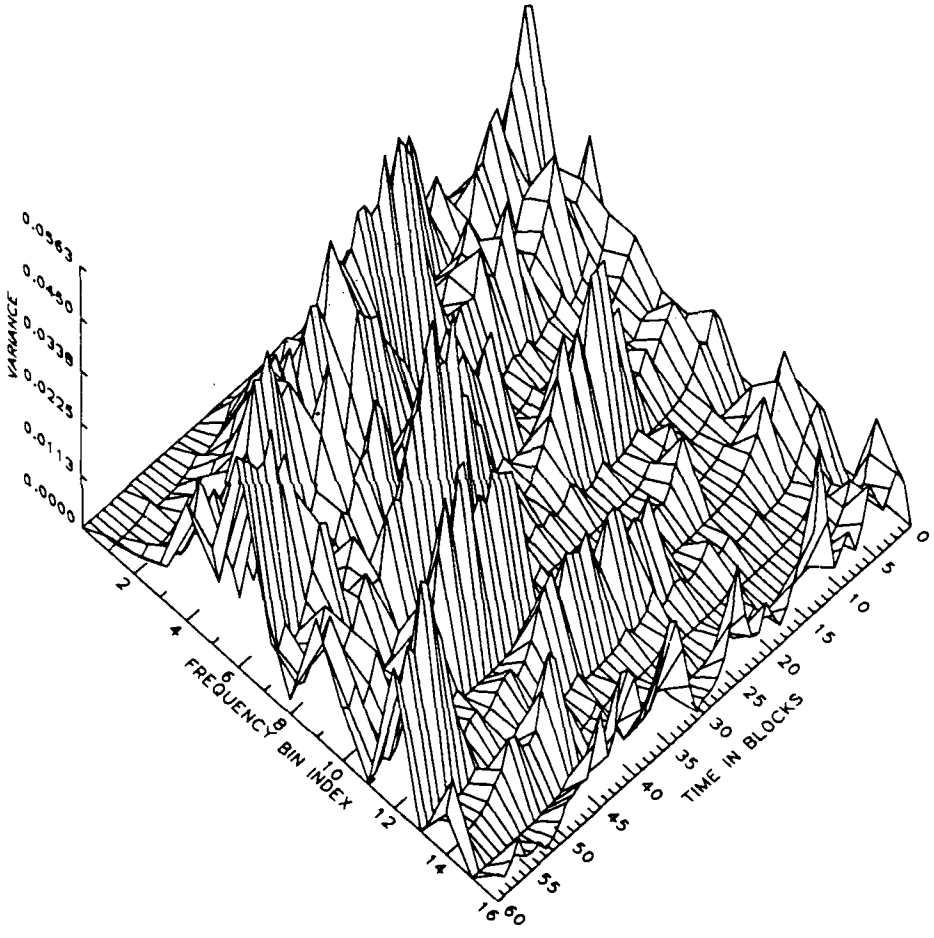


Figure 3.8: Monte Carlo Variance on the Estimated *A Priori* Probability Structure (SNR=6dB).

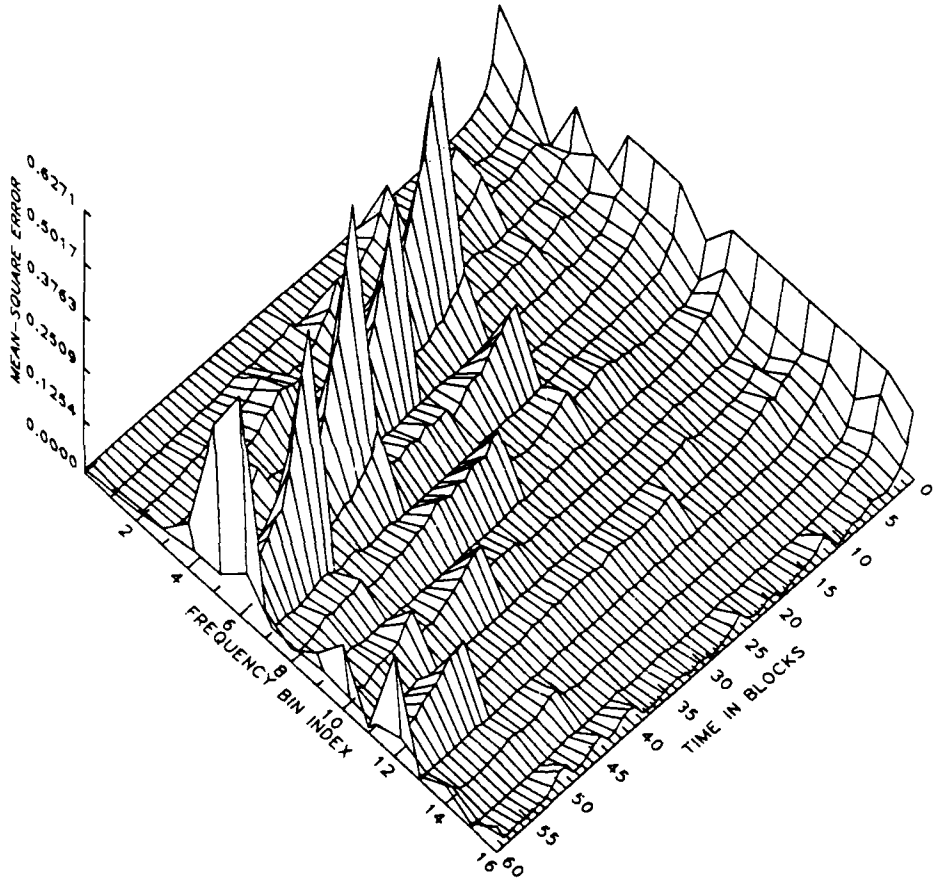


Figure 3.9: Mean-square Error for the Estimated *A Priori* Probability Structure (SNR=6dB).

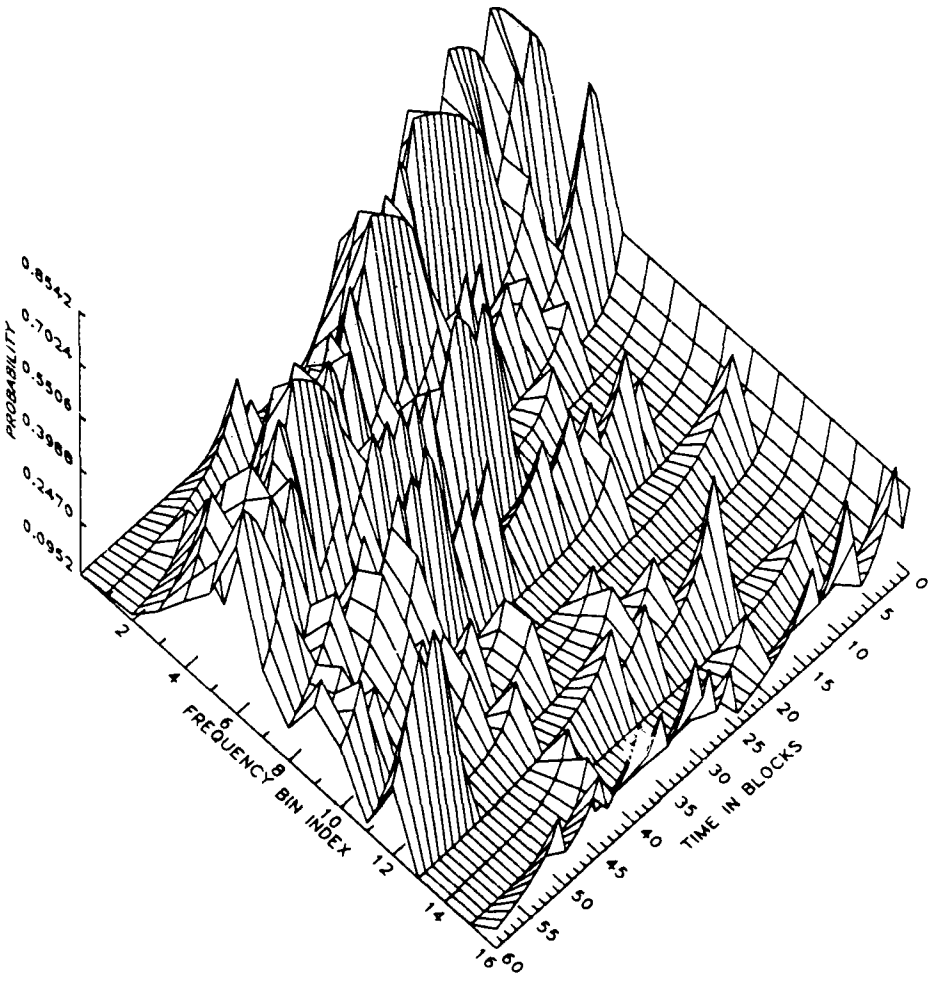


Figure 3.10: Estimated *A Priori* Probability Structure (SNR=12dB).

and used in the GLRT. As can be seen from the plots in Figures 3.10 and 3.11, the estimator is still performing quite well at a signal-to-noise-ratio of 12 dB. However, upon examination of the variance plot, we see that the estimator variance for this case actually has a lower absolute maximum than the variance for case one. Nevertheless, the overall variance seems to be about the same as in case one. The mean-square error for the case two estimator is slightly greater than that for case one.

When the signal-to noise-ratio is 6dB, Figures 3.14 and 3.15 show a significant false alarm rate in FFT bin 1. Since Figure 3.16 shows a high estimator variance in bin 1, it is likely that the false alarm condition shown in Figure 3.14 did not occur on all Monte

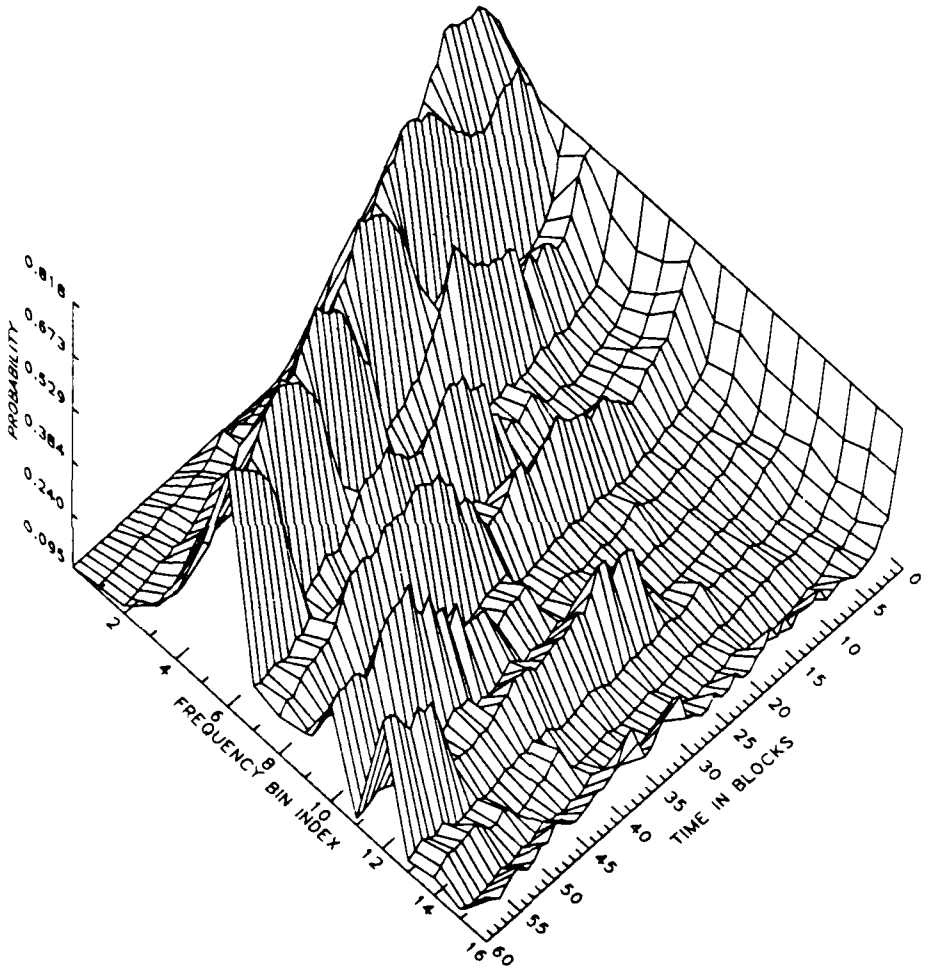


Figure 3.11: Monte Carlo Averages for the Estimated *A Priori* Probability Structure (SNR=12dB).

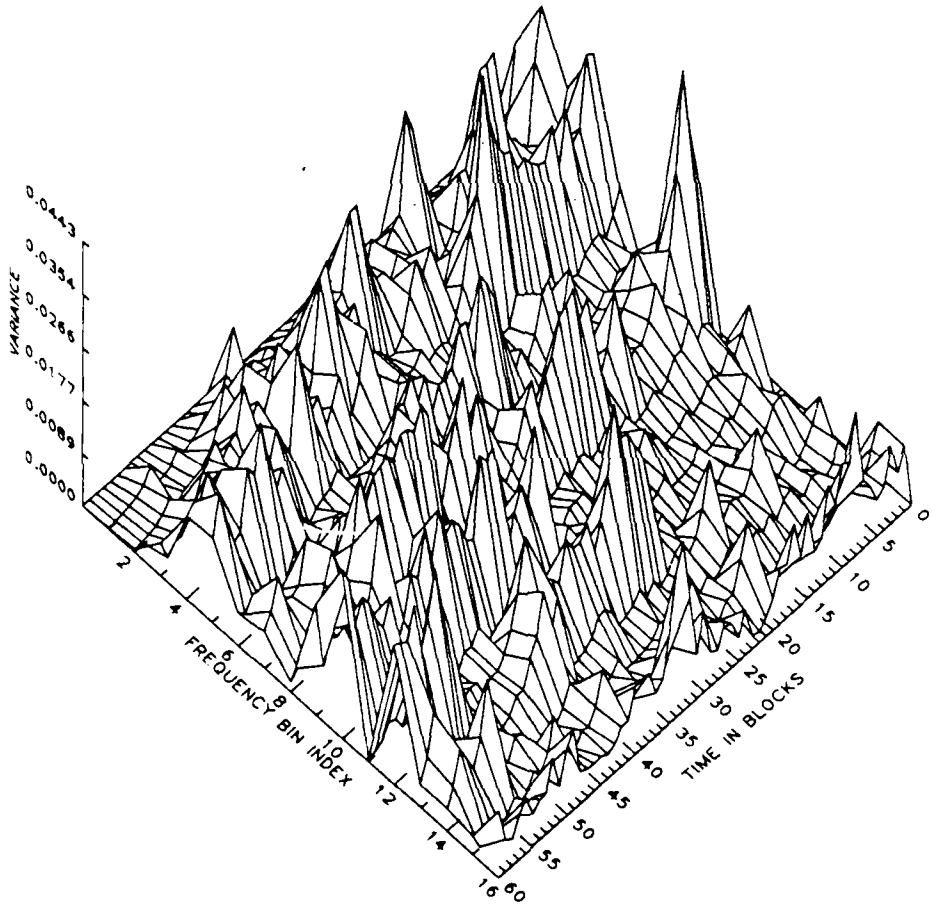


Figure 3.12: Monte Carlo Variance on the Estimated *A Priori* Probability Structure (SNR=12dB).

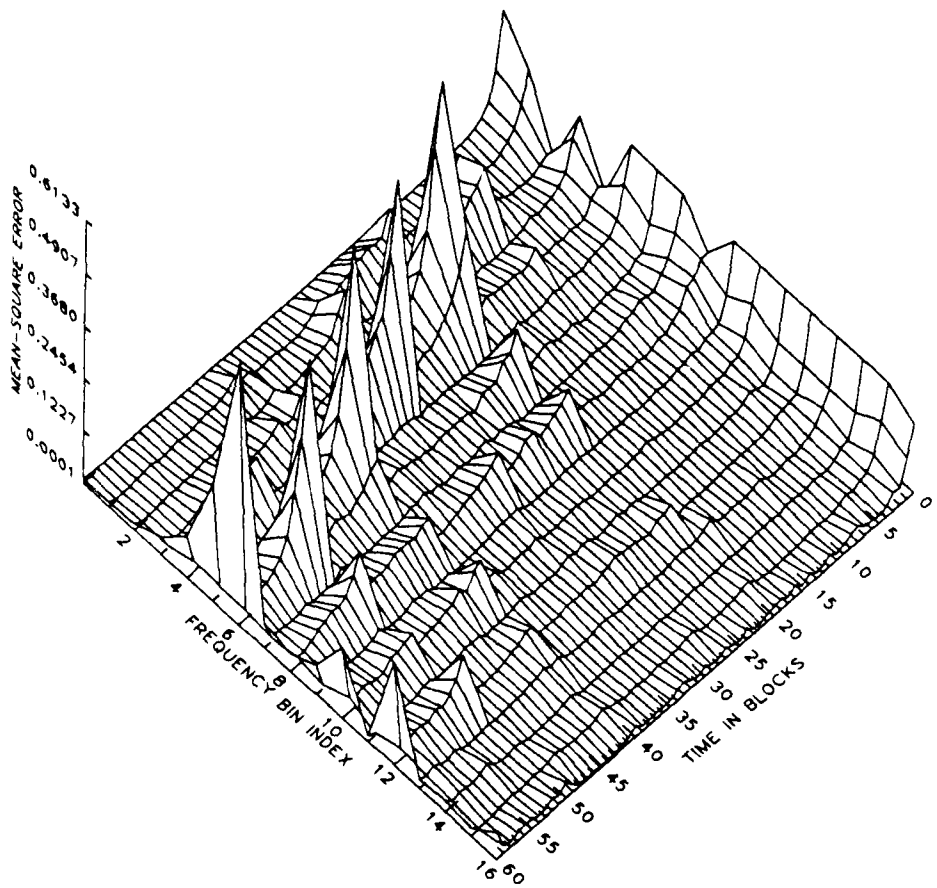


Figure 3.13: Mean-square Error for the Estimated *A Priori* Probability Structure (SNR=12dB).

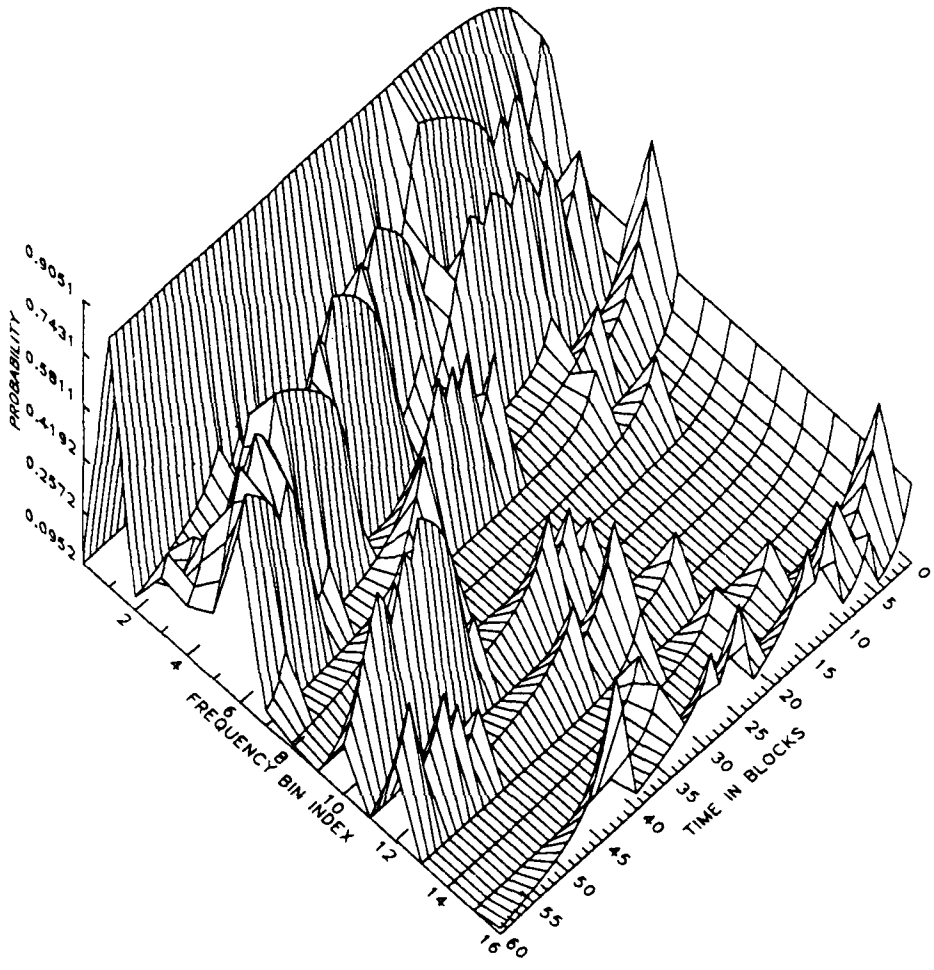


Figure 3.14: Estimated *A Priori* Probability Structure (SNR=6dB).

Carlo simulations. This high false alarm probability in bin 1 can be explained as a runaway condition probably due to the inaccurate estimation of the bin 1 signal amplitude. With the exception of bin 1, however, the estimator variance shown in Figure 3.16 is only somewhat larger than the variance for the 12 dB SNR simulations. Even given the rather erroneous estimation of the probabilities, the mean-square error plot given in Figure 3.17 is not dominated by the problems in bin 1, but rather shows the characteristic peaks along the high-probability trace.

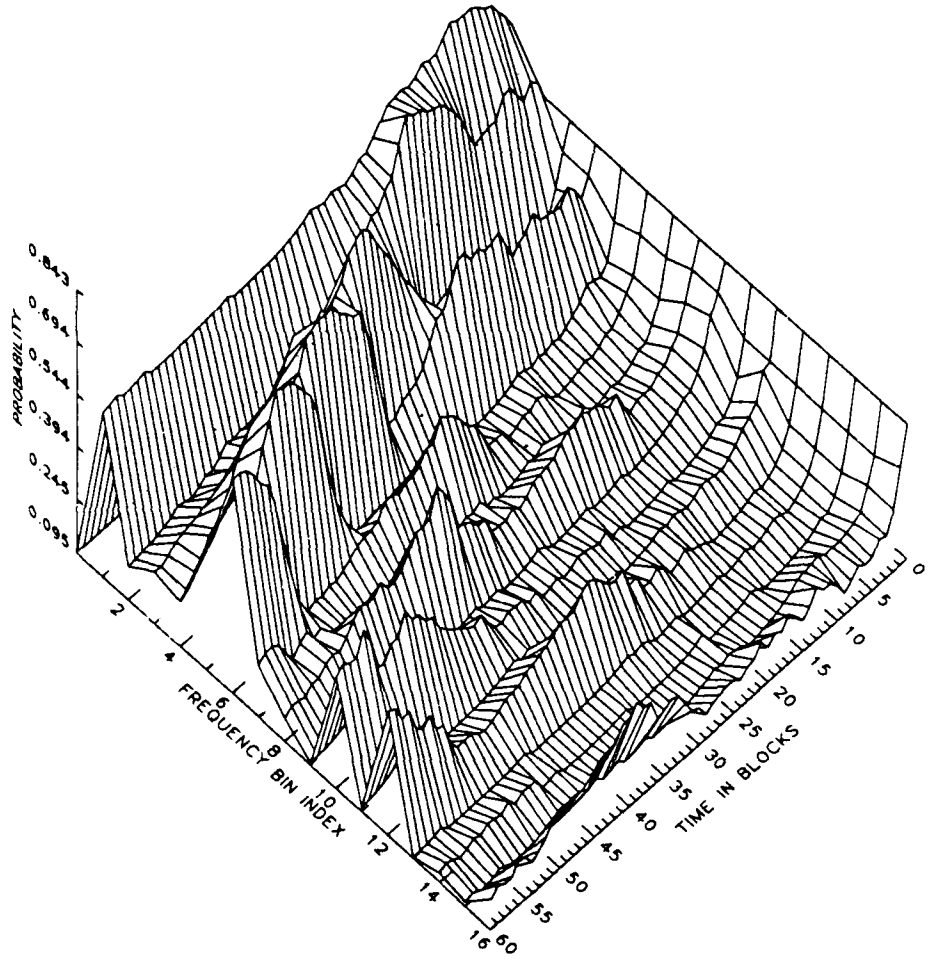


Figure 3.15: Monte Carlo Averages for the Estimated *A Priori* Probability Structure (SNR=6dB).

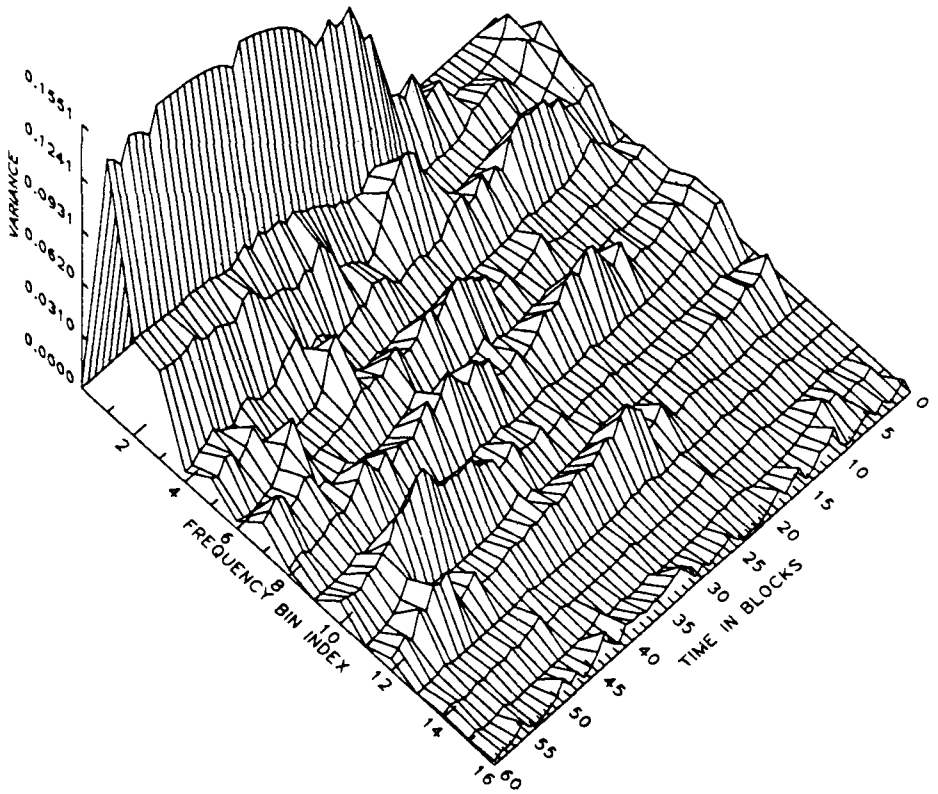


Figure 3.16: Monte Carlo Variance on the Estimated *A Priori* Probability Structure (SNR=6dB).

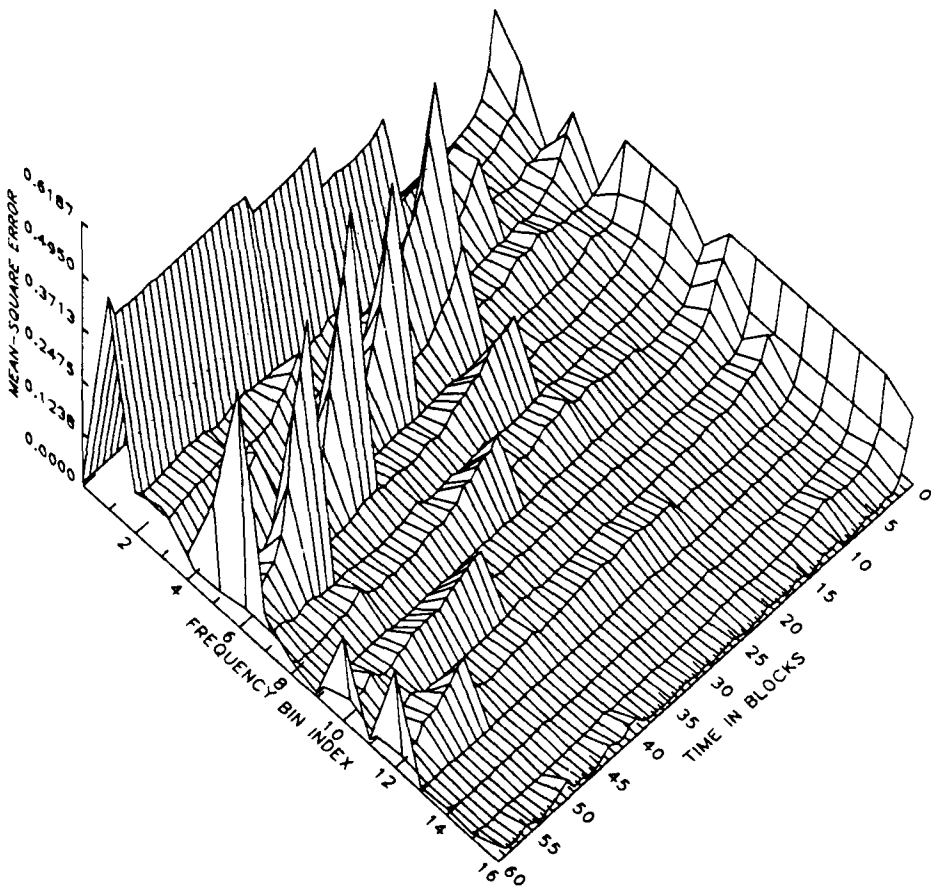


Figure 3.17: Mean-square Error for the Estimated *A Priori* Probability Structure (SNR=6dB).

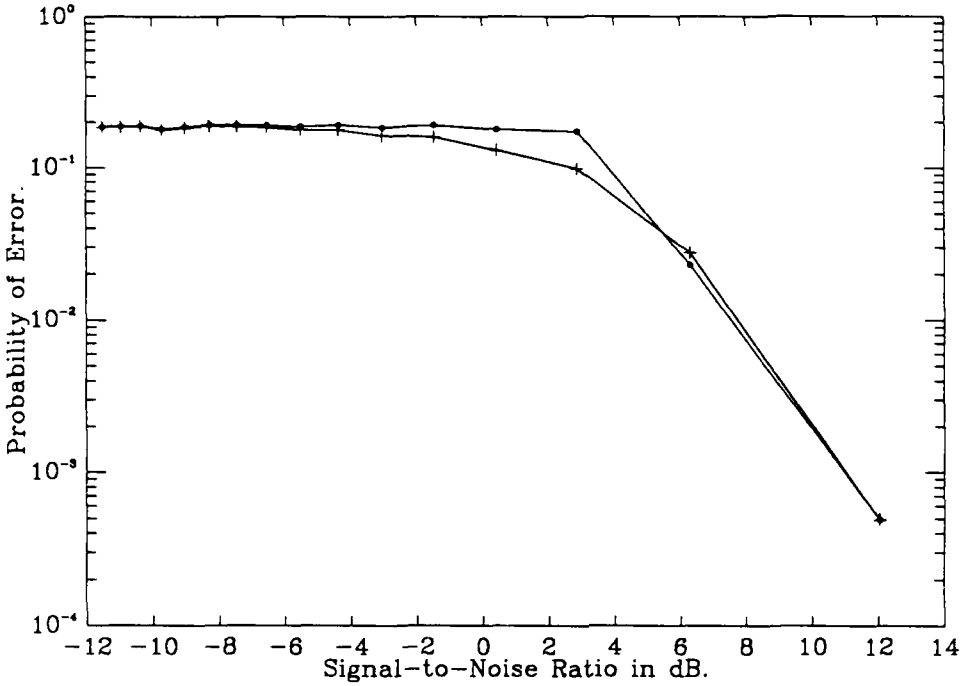


Figure 3.18: Receiver Operating Characteristics - Known Amplitude, Estimated *A Priori* Probabilities.

3.3.2 Receiver Operating Characteristics

Whenever detection performance is evaluated, the common performance evaluation is done using the error probabilities at various signal-to-noise ratios. Plots generated using such a method are referred to as *receiver operating characteristic* (ROC) curves. In Figure 3.18 we give these curves for the case where only the *a priori* probabilities are estimated for the various FFT bins. The line marked by dots is the ROC curve for the empirical Bayes procedure, and the line marked by the crosses is the ROC curve for the true Bayes detection procedure. An unexpected feature of these curves is that they cross at approximately the 6 dB point and are identical at the 12 dB point. The difference at the 6 dB point would be approximately .02 percent, a very small advantage for the empirical Bayes procedure. From Robbins work [22], we do expect that the empirical Bayes and the true Bayes approaches should converge at high signal-to-noise ratios. This is obviously the case here.

Similarly, the curves shown in Figure 3.19 are given for the case where the signal am-

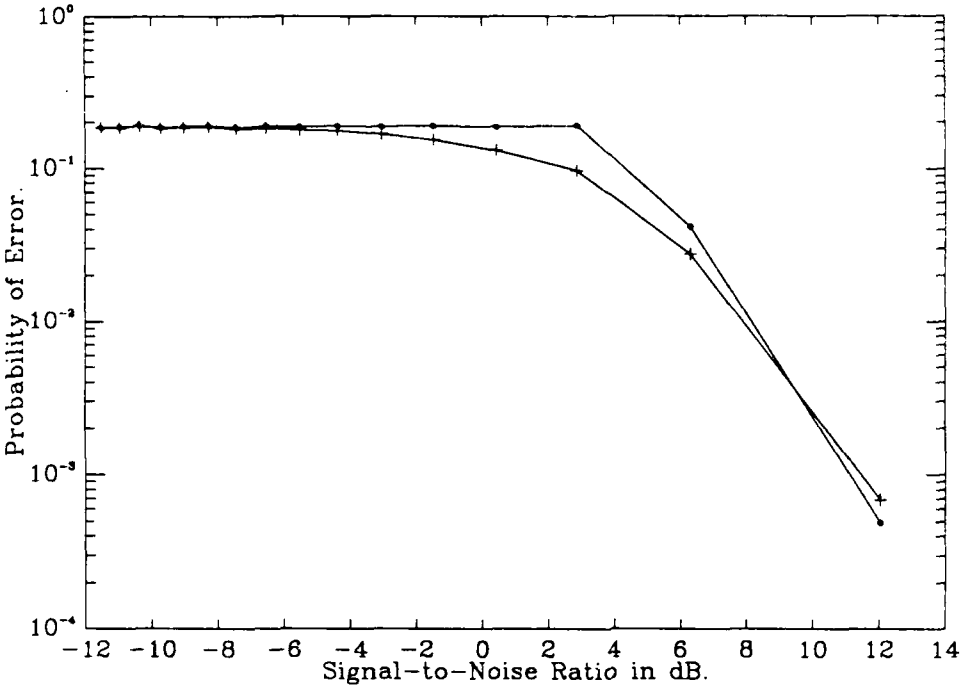


Figure 3.19: Receiver Operating Characteristics - Estimated Amplitude and *A Priori* Probabilities.

plitudes is estimated as well as the *a priori* probabilities. In this plot, the error probability for the empirical Bayes procedure is slightly lower than the true Bayes only at the 12 dB point. Again, this is a very slight difference and is possibly an anomaly in the data which would be remedied if a larger number of Monte Carlo trials were conducted and averaged. From these two graphs then, we see that when the signal bin amplitude must be computed no serious degradation of the algorithm performance occurs.

Bibliography

[1] W. S. Burdic. *Underwater Acoustic System Analysis.* Prentice-Hall, 1984.

[2] J. A. Cadzow, Y. S. Kim, D. C. Shiue, Y. Sun, and G. Xu. Resolution of coherent signals using a linear array. In *Proceedings of IEEE ICASSP, 1987*, Dallas, Texas, April 1987.

[3] T. A. C. M. Claasen and W. F. G. Mecklenbrauker. The Wigner distribution - a tool for time-frequency signal analysis, Part III: relations with other time-frequency signal transformation. *Philips Journal of Research*, 35:372–389, 1980.

[4] T. A. C. M. Claasen and W. F. G. Mecklenbrauker. The Wigner distribution - a tool for time-frequency signal analysis, Part II: discrete-time signals. *Philips Journal of Research*, 35:276–300, 1980.

[5] T. A. C. M. Claasen and W. F. G. Mecklenbrauker. The Wigner distribution - a tool for time-frequency signal analysis, Part I: continuous-time signals. *Philips Journal of Research*, 35:217–250, 1980.

[6] L. D. Davisson and S. C. Schwartz. Analysis of a decision-directed receiver with unknown priors. *IEEE Trans. Inform. Theory*, IT-16(3):270-276, May 1970.

[7] J.L. Doob. *Stochastic Processes.* John Wiley and Sons, Inc., 1953.

[8] W. D. Gregg and J. C. Hancock. An optimum decision-directed scheme for Gaussian mixtures. *IEEE Trans. Inform. Theory*, IT-14(3):451-461, May 1968.

[9] J. K. Hammond and R. F. Harrison. Wigner-ville and evolutionary spectra for covariance equivalent nonstaionary random processes. In *Proceedings of IEEE ICASSP*,

1985, Tampa, Florida, April 1985.

[10] A. G. Jaffer and S. C. Gupta. Recursive Bayesian estimation with uncertain observation. *IEEE Trans. Inform. Theory*, IT-17(5):614–616, September 1971.

[11] D. L. Jones and T. W. Parks. A high resolution data-adaptive time-frequency representation. In *Proceedings of IEEE ICASSP, 1987*, Dallas, Texas, April 1987.

[12] A. Katopis and S. C. Schwartz. Decision-directed learning using stochastic approximation. In *Proc. 3rd Ann. Pittsburgh Conf. Modeling and Simulation*, pages 473–481, Univ. Pittsburgh, Pittsburgh, PA, April 1972.

[13] D. Kazakos and L. D. Davisson. An improved decision-directed detector. *IEEE Trans. Inform. Theory*, IT-26(1):113–116, January 1980.

[14] N. E. Nahi. Optimal recursive estimation with uncertain observation. *IEEE Trans. Inform. Theory*, IT-15(4):457–462, July 1969.

[15] J. Neyman. Two breakthroughs in the theory of statistical decision making. *Rev. Inst. Internat. Statist.*, 30:11–27, 1962.

[16] J. H. Painter and S. C. Gupta. Recursive ideal observer detection of known M-ary signals in multiplicative and additive noise. *IEEE Trans. Commun.*, COM-21(8):948–953, August 1973.

[17] E. A. Patrick and J. P. Costello. Asymptotic probability of error using two decision-directed estimators for two unknown mean vectors. *IEEE Trans. Inform. Theory*, IT-14(1):160–162, January 1968.

[18] E. A. Patrick, J. P. Costello, and F. C. Monds. Decision-directed estimation of a two-class decision boundary. *IEEE Trans. Computers*, C-19(3):197–205, March 1970.

[19] F. Peyrin and R. Prost. A unified definition for the discrete-time, discrete-frequency, and discrete-time/frequency Wigner distributions. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-34(4):858–867, August 1986.

[20] J. G. Proakis, P. R. Drouilhet, Jr., and R. Price. Performance of coherent detection systems using decision-directed channel measurement. *IEEE Trans. Commun. Syst.*, CS-12(1):54–63, March 1964.

[21] H. Robbins. Asymptotically subminimax solutions of compound statistical decision problems. In *Proc. Second Berkley Symposium on Mathematical Statistics and Probability*, pages 131–148, University of California Press, 1950.

[22] H. Robbins. An empirical Bayes approach to statistics. In *Proc. Third Berkley Symposium on Mathematical Statistics and Probability*, pages 157–164, University of California Press, 1955.

[23] Y. Rosen and B. Porat. Arma parameter estimation based on sample covariances for missing data. In *Proceedings of IEEE ICASSP, 1986*, Tokyo, Japan, April 1986.

[24] E. Samuel. Asymptotic solutions of sequential compound decision problem. *Ann. Math. Statist.*, 34:1079–1094, 1963.

[25] R. O. Schmidt. *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation.* PhD thesis, Stanford University, 1981.

[26] S. C. Schwartz and L. D. Davisson. Analysis of a decision-directed receiver for a Markov sequence with unknown transition probabilities. *Problemy Peredachi Informatsii*, 6(2):92–86, April-June 1970.

[27] S. C. Schwartz and A. Katopis. Modified stochastic approximation to enhance unsupervised learning. In *Proc. 1977 Conf. Decision and Control*, pages 1067–1079, 1977.

[28] H. J. Scudder, III. Probability of error of some adaptive pattern-recognition machines. *IEEE Trans. Inform Theory*, IT-11(3):363–371, July 1965.

[29] A. Segall. Recursive estimation from discrete-time point processes. *IEEE Trans. Inform. Theory*, IT-22(4):422–431, July 1976.

[30] W. C. Stirling. *Simultaneous Jump Excitation Modeling and System Parameter Estimation.* PhD thesis, Stanford University, 1983.

[31] W. C. Stirling. Simultaneous system identification and decision-directed detection and estimation of jump inputs to linear systems. *IEEE Trans. Automatic Control*, AC-32(1):86–89, January 1987.

[32] W. C. Stirling and M. Morf. A new decision-directed algorithm for nonstationary priors. *IEEE Trans. Automatic Control*, AC-29(10):928–930, October 1984.

[33] W. C. Stirling and A. L. Swindlehurst. Decision-directed multivariate empirical Bayes classification with nonstationary priors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Sept. 1987. (To Appear).

[34] W. C. Stirling and A. L. Swindleurst. *Decision-Directed Detection of Spatio-Temporal Sources.* Final Report TR S101-86.1, Brigham Young University, Provo, Utah, August 1986.

[35] A. L. Swindlehurst and W. C. Stirling. An adaptive empirical Bayes decision-directed detector. In *Proceedings of 20th Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, November 1986.

[36] A. D. Whalen. *Detection of Signals in Noise.* Academic Press, 1971.

Appendix A

Key Results from Martingale Theory

The most common applications of stochastic process theory in engineering deals with linear estimation (e.g., Wiener and Kalman filters), which rely heavily on the notion of uncorrelation and white noise. For the purposes of the analysis in this report, however, the estimation problems are nonlinear, and these notions are replaced as fundamental concepts by the martingale property. Consequently, it may be useful to provide some background on martingale theory which will render the results of this paper more understandable.

A.1 Background

Definition: A *σ-field* is a collection of subsets, $\mathcal{A}$, of an event space, Ω, such that

(a) For every $A \in \mathcal{A}$, we have the complement, $\overline{A} \in \mathcal{A}$.

(b) if $A_1, A_2, \ldots, A_n, \ldots$, is a countable sequence of elements of $\mathcal{A}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$.

(c) $\emptyset \in \mathcal{A}$.

The elements of $\mathcal{A}$ are termed *events*; the set Ω is the *sure* event, and event $\emptyset$ is the *impossible* event.

Definition: A *probability space* is a triple, $\{\Omega, \mathcal{A}, P\}$ where Ω is the event space, $\mathcal{A}$ is a σ-field, and P is a probability measure on $\mathcal{A}$, i.e., P satisfies

(a) $P\{A\} \geq 0$, for all $A \in \mathcal{A}$.

(b) $P\{\Omega\} = 1$

(c) If $\{A_n\}$ is a countable collection of disjoint events, then $P\{\cup_{n=1}^{\infty} A_n\} = \sum_{n=1}^{\infty} P\{A_n\}$

Definition: A *random variable* X is a real-valued function whose domain is Ω and which is $\mathcal{A}$-measurable, i.e., for every real number x, the set $\{\omega \in \Omega | X(\omega) \leq x\} \in \mathcal{A}$.

Definition: The σ-field *generated* by the set of random variables $\{X_t, t \in T\}$ is the smallest σ-field with respect to which each element of $\{X_t,\ t \in T\}$ is measurable. This σ-field is denoted $\sigma\{X_t,\ t \in T\}$. The physical meaning of the σ-field $\mathcal{B}_T = \sigma\{X_t,\ t \in T\}$ is that $\mathcal{B}_T$ represents all of the information contained in or derived from the collection of random variables $\{X_t\ t \in T\}$.

Definition: Let X be a random variable on $\{\Omega, \mathcal{A}, P\}$ with finite expectation, and let $\mathcal{B}$ be a sub σ-field of $\mathcal{A}$. The *conditional expectation* of X with respect to $\mathcal{B}$, denoted $E^{\mathcal{B}}X$ or $E(X|\mathcal{B})$ is a $\mathcal{B}$-measurable random variable, uniquely determined except over a $\mathcal{B}$-measurable event of probability zero, which satisfies

$$\int_B X dP = \int_B E^{\mathcal{B}} X dP$$

for all $B \in \mathcal{B}$. Although it will not be proven here, it is a standard result of mathematical probability theory (a consequence of the Radon-Nikodym theorem) that the existence and uniqueness (up to $\mathcal{B}$-measurable events of probability zero) of such a function $E^{\mathcal{B}}X$ is assured. In the sequel, conditions that hold except possible on a set of probability zero will be said to hold almost surely, or a.s. Some basic properties of conditional expectations may be summarized as

(a) $E^{\mathcal{B}}X \geq 0$ if $X \geq 0$ a.s.; $E^{\mathcal{B}}X = 0$ if $X = 0$ a.s.

(b) $E(E^{\mathcal{B}}X) = E(X)$ a.s.; $E^{\mathcal{B}}(1) = 1$ a.s.

(c) $E^{\mathcal{B}}(cX) = cE^{\mathcal{B}}X$ if $c \in \Re$.

(d) $E^{\mathcal{B}}(X_1 + X_2) = E^{\mathcal{B}}X_1 + E^{\mathcal{B}}X_2$ if both expectations exist.

(e) If X is $\mathcal{B}$-measurable, then $E^{\mathcal{B}}X = X$ a.s. and, more generally, for every random variable Y on $\{\Omega, \mathcal{B}, P\}$, we have $E^{\mathcal{B}}(XY) = X \cdot E^{\mathcal{B}}Y$.

(f) If $\mathcal{B}_1 \in \mathcal{B}_2$, then $E^{\mathcal{B}_1}(E^{\mathcal{B}_2}X) = E^{\mathcal{B}_1}X = E^{\mathcal{B}_2}(E^{\mathcal{B}_1}X)$.

If a random variable X is measurable with respect to a σ-field $\mathcal{B}$, this fact is denoted by the notation $X \in \mathcal{B}$.

Definition: Let $\{\mathcal{B}_t,\ t = 0, 1, \ldots\}$ be an increasing family of σ-fields and let $\{X_t,\ t = 0, 1, \ldots,\}$ be a set of random variables such that $X_t \in \mathcal{B}_t$. Then $\{X\}$ is said to be *adapted* to $\{\mathcal{B}\}$. The family of random variables $\{X\}$ is called a *martingale* with respect to $\{\mathcal{B}\}$ if

(a) $X_t \in \mathcal{B}_t$

(b) $E^{\mathcal{B}_s}X_t = X_s$ for $s < t$.

A *sub-martingale* is defined as above with the exception that 2) is replaced by $E^{\mathcal{B}_s}X_t \geq X_s$.

A *martingale difference*, or MD, process $\{x\}$ is formed from a martingale process $\{X\}$ by defining $x_t = X_t - X_{t-1}$. Thus,

$$E^{\mathcal{B}_{t-1}}x_t = E^{\mathcal{B}_{t-1}}X_t - X_{t-1} = X_{t-1} - X_{t-1} = 0.$$

It is useful to characterize the MD property as being intermediate between the properties of independence and uncorrelation, since every two independent random variables have the MD property with respect to each other and, in turn, every two random variables that have the MD property are uncorrelated.

Definition: For a process (possibly vector-valued) $\{\mathbf{X}\}$ adapted to $\{\mathcal{B}\}$, the *conditional variance* $(\mathbf{X}, \mathbf{X})_t$ sequence is

$$(\mathbf{X}, \mathbf{X})_t = E^{\mathcal{B}_{t-1}}\mathbf{X}(t)\mathbf{X}^T(t).$$

For any two processes $\{\mathbf{X}\}$ and $\{\mathbf{Y}\}$ the *conditional covariance* $(\mathbf{X}, \mathbf{Y})_t$ is

$$(\mathbf{X}, \mathbf{Y})_t = E^{\mathcal{B}_{t-1}}\mathbf{X}(t)\mathbf{Y}^T(t).$$

Definition: The sequence $\{b\}$ of random variables is said to be $\{\mathcal{B}\}$-*predictable* if $b_t \in \mathcal{B}_{t-1}$ for all $t = 0, 1, \ldots$. If $\{b\}$ is $\{\mathcal{B}\}$-predictable and $\{v\}$ is a MD with respect

to $\{\mathcal{B}\}$, we define the process $\{b \cdot v\}$ by $b_t \cdot v_t = b_t v_t$, the product. Then $\{b \cdot v\}$ is a MD sequence with respect to $\{\mathcal{B}\}$. This sequence is termed the *MD transform* of $\{v\}$ by $\{b\}$.

Definition: For any sequence of random variables $\{y_t,\ t = 0, 1, \ldots\}$, let $\mathcal{F}_t = \sigma\{y_s, s = 0, 1, \ldots t\}$. Then the sequence $\nu_t = y_t - E^{\mathcal{F}_{t-1}} y_t$ is the *general innovations* of the process $\{y\}$.

The development of the discrete-time point process estimator used in this report relies two basic properties that are central to nonlinear estimation theory. These two properties are: 1) the innovations theorem, and 2) the representation theorem. These results are stated without proof for the discrete-time case.

Innovations Theorem: The general innovations process is a martingale that is equivalent to the observed process, i.e.,

$$\mathcal{F}_t = \sigma\{y_s, s \leq t\} = \sigma\{\nu_s, s \leq t\}.$$

Representation Theorem: Every MD sequence $\{w\}$ with respect to the $\{\mathcal{F}\}$ can be represented as a MD transform of the innovations $\{\nu\}$, i.e.,

$$w_t = b_t \nu_t$$

where $\{b\}$ is an $\{\mathcal{F}\}$ predictable process.

A.2 Discrete-Time Point Processes

A discrete-time point process (DTPP) $\{N(t), t = 0, 1, \ldots\}$ is a binary $\{0,\ 1\}$ sequence describing the occurrences of some type of (possibly vector-valued) event. Thus, $N(t) = 1$ means that the event occurs at time t, and $N(t) = 0$ means that there is no occurrence of the event at time t. The simplest example is when $\{N\}$ is a Bernoulli sequence, i.e., a sequence of *independent* random variables with $P\{N(t) = 1\} = 1 - P\{N(t) = 0\} = \lambda(t)$. The quantity $\lambda(t)$ is the *rate* of $N(t)$, and may in general be time-varying. In many applications the occurrences will not be mutually

independent, but the probability of occurrence at a given time will be affected by previous occurrences and perhaps by some other related process, $\mathbf{x}(t)$, in which case the rate of the process will be affected by the past history of $\{\mathbf{x}\}$ as well as by $\{N\}$.

A.2.1 Doob Decomposition

We present a fundamental theorem due to Doob [7].

Doob Decomposition: For an arbitrary sequence $\{y\}$ adapted to a family $\{\mathcal{B}\}$ of σ-fields, define

$$\lambda(t) = E^{\mathcal{B}_{t-1}} y(t)$$

and

$$w(t) = y(t) - E^{\mathcal{B}_{t-1}} y(t).$$

Then the following properties hold:

$$y(t) = \lambda(t) + w(t)$$

(a) $\{\lambda\}$ is $\{\mathcal{B}\}$-predictable and $\{w\}$ is a MD sequence with respect to $\{\mathcal{B}\}$;

(b) the above decomposition is unique;

(c) if $\{y\}$ is a $\{\mathcal{B}\}$-submartingale difference (subMD) sequence, i.e., if

$$E^{\mathcal{B}_{t-1}} y(t) \geq 0,$$

then $\lambda(t)$ is positive for all t.

Proof: Properties 1) and 3) are trivial. To prove 2), suppose

$$y(t) = \lambda'(t) + w'(t)$$

where $\{\lambda'\}$ is $\{\mathcal{B}\}$-predictable and $\{w'\}$ is a $\{\mathcal{B}\}$ MD sequence. Then

$$0 = E^{\mathcal{B}_{t-1}} w'(t) = E^{\mathcal{B}_{t-1}} y(t) - \lambda'(t) = \lambda(t) - \lambda'(t)$$

which proves 2).

For point processes,

$$N(t) = \lambda(t) + w(t)$$

is *the* Doob decomposition of $\{N\}$ with respect to $\{\mathcal{B}\}$.

A.2.2 Estimation from Discrete-Time Point Processes

It is useful to outline the development of the discrete-time point process estimator used in this analysis. This development follows [29]. Suppose the rate of the DTPP $\{N\}$ $(\boldsymbol{\lambda}(t))$ may be characterized as a finite-state Markov chain, with states $\rho_1(t), \ldots, \rho_r(t)$. Define the vector $\mathbf{x}(t)$ with element

$$x_i(t) = \begin{cases} 1 & \text{if } \boldsymbol{\lambda}(t) = \rho_i(t) \\ 0 & \text{otherwise} \end{cases},$$

and the probability transition matrix $\mathbf{Q}(t)$ with elements

$$q_{ij}(t) = P\{x_j(t+1) = 1 | x_i(t) = 1\}.$$

Define the σ-field

$$\mathcal{B}_t = \sigma\{N(s), s \leq t,\ \mathbf{x}(s), s \leq t+1\}.$$

The r-dimensional vector process $\mathbf{x}(t)$ trivially obeys the relation

$$\mathbf{x}(t+1) = E^{\mathcal{B}_{t-1}}\mathbf{x}(t+1) + [\mathbf{x}(t+1) - E^{\mathcal{B}_{t-1}}\mathbf{x}(t+1).$$

It can easily be seen that the process $E^{\mathcal{B}_{t-1}}\mathbf{x}(t+1)$ is $\{\mathcal{B}\}$-predictable, and that

$$\mathbf{u}(t) = \mathbf{x}(t+1) - E^{\mathcal{B}_{t-1}}\mathbf{x}(t+1)$$

is a $\{\mathcal{B}\}$-MD sequence. Thus,

$$\mathbf{x}(t+1) = E^{\mathcal{B}_{t-1}}\mathbf{x}(t+1) + \mathbf{u}(t) \tag{A.1}$$

is analogous to the classical signal-plus-noise model. It can be seen from construction, however, that this decomposition is fundamentally different from the classical situation, and that the process $\{\mathbf{u}\}$ is not an independent process.

We may evaluate $E^{\mathcal{B}_{t-1}}\mathbf{x}(t+1)$ for the Markov chain model as follows. Since $\mathbf{x}(t)$ completely characterizes the expected behavior of $\mathbf{x}(t+1)$ (due to the Markov structure) we have

$$E^{\mathcal{B}_{t-1}}x_i(t+1) = E[x_i(t+1)|\mathbf{x}(t)] = P\{x_i(t+1) = 1|\mathbf{x}(t)\} = \sum_{j=1}^{r} q_{ji}(t)x_j(t).$$

Thus,

$$E^{\mathcal{B}_{t-1}}\mathbf{x}(t+1) = \mathbf{Q}^T(t)\mathbf{x}(t), \tag{A.2}$$

and we have the representation

$$\mathbf{x}(t+1) = \mathbf{Q}^T(t)\mathbf{x}(t) + \mathbf{u}(t). \tag{A.3}$$

We may define

$$w(t) = N(t) - \lambda(t)$$

where $\lambda(t) = E^{\mathcal{B}_{t-1}}N(t)$. The process $\{w\}$ is a $\{\mathcal{B}\}$ MD process. Under the Markov model, the components of $\mathbf{x}(t)$ are

$$x_i(t) = \begin{cases} 1 & \text{if } \lambda(t) = \rho_i(t) \\ 0 & \text{otherwise} \end{cases}.$$

Thus, we may write

$$N(t) = \boldsymbol{\rho}^T(t)\mathbf{x}(t) + w(t) \tag{A.4}$$

where

$$\boldsymbol{\rho}(t) = \begin{bmatrix} \rho_1^T(t) \\ \cdot \\ \cdot \\ \cdot \\ \rho_r^T(t) \end{bmatrix}.$$

We are interested in the conditional expectation of $\mathbf{x}(t+1)$ given $\mathcal{F}_t$. To obtain this representation, we first form the process

$$\boldsymbol{\mu}(t) = E^{\mathcal{F}_t}\mathbf{x}(t+1) - E^{\mathcal{F}_{t-1}}E^{\mathcal{B}_{t-1}}\mathbf{x}(t+1),$$

and note that this quantity is a $\{\mathcal{F}\}$-MD. Since $\mathcal{F}_{t-1} \subset \mathcal{B}_{t-1}$ and using (A.2) we obtain

$$\boldsymbol{\mu}(t) = E^{\mathcal{F}_t}\mathbf{x}(t+1) - \mathbf{Q}^T(t)E^{\mathcal{F}_{t-1}}\mathbf{x}(t). \tag{A.5}$$

We also note that the process

$$\nu(t) = N(t) - E^{\mathcal{F}_{t-1}}E^{\mathcal{B}_{t-1}}N(t) = N(t) - \boldsymbol{\rho}^T(t)E^{\mathcal{F}_{t-1}}\mathbf{x}(t)$$

is a $\{\mathcal{F}\}$-MD, where we have used (A.4) and the fact that $\{w\}$ is a $\{\mathcal{B}\}$-MD. The representation theorem states that we can express $\{\boldsymbol{\mu}\}$ in terms of $\{\nu\}$ as

$$\boldsymbol{\mu}(t) = \mathbf{B}(t)\nu(t) \tag{A.6}$$

for some $\{\mathcal{F}\}$-predictable (matrix) sequence $\{\mathbf{B}\}$. The matrix $\mathbf{B}(t)$ can be obtained by computing the conditional covariance process with respect to $\{\mathcal{F}\}$, which implies that

$$(\boldsymbol{\mu}, \nu)_t = \mathbf{B}(t)(\nu, \nu)_t$$

which yields

$$\mathbf{B}(t) = (\boldsymbol{\mu}, \nu)_t[(\nu, \nu)_t]^{-1}. \tag{A.7}$$

Thus, from (A.5), (A.6), and (A.7) we obtain

$$E^{\mathcal{F}_t}\mathbf{x}(t+1) = \mathbf{Q}^T(t)E^{\mathcal{F}_{t-1}}\mathbf{x}(t) + (\boldsymbol{\mu}, \nu)_t[(\nu, \nu)_t]^{-1}\nu(t),$$

or, using the notation $\hat{y}(t|s) = E^{\mathcal{F}_s}y(t)$, for any process y, we obtain

$$\hat{\mathbf{x}}(t+1|t) = \mathbf{Q}^T(t)\hat{\mathbf{x}}(t|t-1) + (\boldsymbol{\mu}, \nu)_t[(\nu, \nu)_t]^{-1}\nu(t). \tag{A.8}$$

Calculation of Conditional Covariance

The covariance matrix $(\boldsymbol{\mu}, \nu)_t$ may be expressed as

$$E^{\mathcal{F}_{t-1}}[\boldsymbol{\mu}(t)\nu(t)] = E^{\mathcal{F}_{t-1}}\left\{E^{\mathcal{F}_t}[\mathbf{x}(t+1)]\nu(t) - E^{\mathcal{F}_{t-1}}[\mathbf{x}(t+1)\nu(t)]\right\}$$

and since $E^{\mathcal{F}_{t-1}}\mathbf{x}(t+1) \in \mathcal{F}_{t-1}$ and $\nu(t)$ is a $\{\mathcal{F}\}$-MD, the second term of this expression on the right-hand side vanishes. Thus, we may write

$$\begin{aligned}
E^{\mathcal{F}_{t-1}}[\boldsymbol{\mu}(t)\nu(t)] &= E^{\mathcal{F}_{t-1}}E^{\mathcal{F}_t}[\mathbf{x}(t+1)\nu(t)] - \mathbf{Q}^T(t)E^{\mathcal{F}_{t-1}}[\mathbf{x}(t)\nu(t)] \\
&= E^{\mathcal{F}_{t-1}}\left\{[\mathbf{Q}^T(t)\mathbf{x}(t) + \mathbf{u}(t)][\mathbf{x}^T(t)\boldsymbol{\rho}(t) + w(t) - \hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t)]\right\} \\
&= E^{\mathcal{F}_{t-1}}\left\{\mathbf{Q}^T(t)\mathbf{x}(t)\mathbf{x}^T(t)\boldsymbol{\rho}(t) + \mathbf{Q}^T(t)\mathbf{x}(t)w(t) - \mathbf{Q}^T(t)\mathbf{x}(t)\hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t)\right. \\
&\qquad \left. -\mathbf{u}(t)\mathbf{x}^T(t)\boldsymbol{\rho}(t) + \mathbf{u}(t)w(t) - \mathbf{u}(t)\hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t)\right\}
\end{aligned}$$

where we have used (A.4), (2.20), and (2.15). Noting that $\mathbf{x}(t) \in \mathcal{B}_{t-1}$, $\hat{\mathbf{x}}(t|t-1)$, and $\mathbf{u}(t)$ and $w(t)$ are $\{\mathcal{B}\}$-MD processes, this expression simplifies to

$$E^{\mathcal{F}_{t-1}}[\boldsymbol{\mu}(t)\nu(t)] = E^{\mathcal{F}_{t-1}}\left\{\mathbf{Q}^T(t)[\mathbf{x}(t)\mathbf{x}^T(t) - \mathbf{x}(t)\hat{\mathbf{x}}^T(t|t-1)]\boldsymbol{\rho}(t) + \mathbf{u}(t)w(t)\right\} \tag{A.9}$$

Solving for $E^{\mathcal{F}_{t-1}}[\mathbf{u}(t)w(t)]$ yields

$$E^{\mathcal{F}_{t-1}}[\mathbf{u}(t)w(t)] = E^{\mathcal{F}_{t-1}}\left\{[\mathbf{x}(t+1) - \mathbf{Q}(t)^T\mathbf{x}(t)][N(t) - \hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t)]\right\}.$$

Since $\mathcal{F}_{t-1} \subseteq \mathcal{B}_{t-1}$ we may write

$$\begin{aligned} E^{\mathcal{F}_{t-1}}[\mathbf{u}(t)w(t)] &= E^{\mathcal{F}_{t-1}}E^{\mathcal{B}_{t-1}}\left\{\mathbf{x}(t+1)N(t) - \mathbf{x}(t+1)\hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t)\right. \\ &\quad \left. -\mathbf{Q}^T(t)\mathbf{x}(t)N(t) + \mathbf{Q}^T(t)\mathbf{x}(t)\hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t)\right\} \end{aligned}$$

where we have multiplied out the cross products. Substituting (A.3) and (A.4) yields

$$\begin{aligned} E^{\mathcal{F}_{t-1}}[\mathbf{u}(t)w(t)] &= E^{\mathcal{F}_{t-1}}E^{\mathcal{B}_{t-1}}\left\{\mathbf{x}(t+1)N(t) - [\mathbf{Q}^T(t)\mathbf{x}(t) + \mathbf{u}(t)]\hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t)\right. \\ &\quad \left. - \mathbf{Q}^T(t)\mathbf{x}(t)[\mathbf{x}^T(t)\boldsymbol{\rho}(t) + w(t)] + \mathbf{Q}^T(t)\mathbf{x}(t)\hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t)\right\} \end{aligned}$$

We note that since $\hat{\mathbf{x}}(t|t-1) \in \mathcal{B}_{t-1}$ and $\mathbf{u}(t)$ is a $\{\mathcal{B}\}$-MD, we have

$$E^{\mathcal{B}_{t-1}}[\mathbf{u}(t)\hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t)] = 0.$$

Similarly, since $\mathbf{x}(t) \in \mathcal{B}_{t-1}$ and $w(t)$ is a $\{\mathcal{B}\}$-MD, we also have

$$E^{\mathcal{B}_{t-1}}[\mathbf{Q}^T(t)\mathbf{x}(t)w(t)] = 0.$$

Thus, after simplification, we have

$$E^{\mathcal{F}_{t-1}}[\mathbf{u}(t)w(t)] = E^{\mathcal{F}_{t-1}}E^{\mathcal{B}_{t-1}}\left\{\mathbf{x}(t+1)N(t) - \mathbf{Q}^T(t)\mathbf{x}(t)\mathbf{x}^T(t)\boldsymbol{\rho}(t)\right\}.$$

Substituting this expression into (A.9) yields, after straightforward manipulation,

$$E^{\mathcal{F}_{t-1}}[\boldsymbol{\mu}(t)\nu(t)] = E^{\mathcal{F}_{t-1}}[\mathbf{x}(t+1)N(t)] - \mathbf{Q}^T(t)\hat{\mathbf{x}}(t|t-1)\hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t). \quad \text{(A.10)}$$

To complete the development we will assume that, given $\mathcal{B}_{t-1}$, the values assumed by $\mathbf{x}(t+1)$ and $N(t)$ are independent and, consequently,

$$E^{\mathcal{B}_{t-1}}[\mathbf{x}(t+1)N(t)] = E^{\mathcal{B}_{t-1}}\mathbf{x}(t)E^{\mathcal{B}_{t-1}}N(t) = \mathbf{Q}^T(t)\mathbf{x}(t)\mathbf{x}^T(t)\boldsymbol{\rho}(t). \quad \text{(A.11)}$$

But

$$E^{\mathcal{F}_{t-1}}[\mathbf{x}(t+1)N(t)] = E^{\mathcal{F}_{t-1}}E^{\mathcal{B}_{t-1}}[\mathbf{x}(t+1)N(t)],$$

therefore,

$$E^{\mathcal{F}_{t-1}}[\boldsymbol{\mu}(t)\nu(t)] = E^{\mathcal{F}_{t-1}}[\mathbf{Q}^T(t)\mathbf{x}(t)\mathbf{x}^T(t)\boldsymbol{\rho}(t)] - \mathbf{Q}^T(t)\hat{\mathbf{x}}(t|t-1)\hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t)$$

but, since $\mathbf{x}(t)$ has one and only one non-zero component, we may write

$$\mathbf{x}(t)\mathbf{x}^T(t) = \text{diag}\,\{\mathbf{x}(t)\}$$

where diag $\{\cdot\}$ denotes a diagonal matrix whose diagonal elements are composed of the elements of the vector argument. Thus,

$$E^{\mathcal{F}_{t-1}}[\boldsymbol{\mu}(t)\nu(t)] = \mathbf{Q}^T(t)\text{diag}\,\{\hat{\mathbf{x}}(t|t-1)\}\boldsymbol{\rho}(t) - \mathbf{Q}^T(t)\hat{\mathbf{x}}(t|t-1)\hat{\mathbf{x}}^T(t|t-1)\boldsymbol{\rho}(t). \quad \text{(A.12)}$$

Calculation of Conditional Variance

The conditional variance $(\nu,\nu)_t$ consists of the quantity

$$E^{\mathcal{F}_{t-1}}[\nu(t)\nu(t)] = \left\{E^{\mathcal{F}_{t-1}}(N(t) - \hat{\lambda}(t|t-1))(N(t) - \hat{\lambda}(t|t-1))\right\}$$

where

$$\hat{\lambda}(t|t-1) = \boldsymbol{\rho}^T(t)\hat{\mathbf{x}}(t|t-1) \quad \text{(A.13)}$$

is the conditional expectation of $\lambda(t)$ given $\mathcal{F}_{t-1}$. Thus, $E^{\mathcal{F}_{t-1}}\nu^2(t)$ may be easily evaluated as

$$\begin{aligned} E^{\mathcal{F}_{t-1}}\nu^2(t) &= E^{\mathcal{F}_{t-1}}(N^2(t) - 2N(t)\hat{\lambda}(t|t-1) + \hat{\lambda}^2(t|t-1)) \\ &= \hat{\lambda}(t|t-1) - \hat{\lambda}^2(t|t-1) \end{aligned}$$

since $N^2(t) = N(t)$ and $\hat{\lambda}(t|t-1) \in \mathcal{F}_{t-1}$.

Appendix B

Envelope-squared Detection

B.1 Envelope of a Narrowband Process

A random signal $n(t)$ is said to be a narrowband noise process if the spectrum of $n(t)$ is zero except for a narrow region about $\omega = 0$ or $\omega = \omega_c$ where ω_c is a carrier frequency. Let us represent the process $n(t)$ as a pair of quadrature components as given in Whalen [36] as follows:

$$n(t) = x(t)\cos(\omega_c t) - y(t)\sin(\omega_c t) \tag{B.1}$$

where $x(t)$ and $y(t)$ are also narrowband processes with power or variance σ^2. We then refer to the *envelope* of the process $n(t)$ as

$$z(t) = [x^2(t) + y^2(t)]^{\frac{1}{2}}. \tag{B.2}$$

Now, define the phase angle as

$$\theta(t) = \arctan\frac{y(t)}{x(t)}. \tag{B.3}$$

The inverse operations are then

$$x(t) = z(t)\cos\theta \qquad y(t) = z(t)\sin\theta,$$

and the Jacobian of the transformation is $z(t)$. The joint probability density function for $z(t)$ and θ is, suppressing the time arguments:

$$p(z,\theta) = \frac{z\exp\frac{-z^2}{2\sigma^2}}{2\pi\sigma^2}. \tag{B.4}$$

Since the phase angle is uniformly distributed in $[0, 2\pi]$, we integrate Equation B.4 to get

$$p(z) = \int_0^{2\pi} p(z, \theta) d\theta = \frac{z}{\sigma^2} \exp\left[-\frac{z^2}{2\sigma^2}\right] \quad z \geq 0. \tag{B.5}$$

This is the well-known Rayleigh density function. However, since we wish to employ the square of the envelope in a detector, we now let $u(t) = z^2(t)$, and the density function becomes

$$p(u) = \frac{1}{2\sigma^2} \exp\left[-\frac{u}{2\sigma^2}\right]. \tag{B.6}$$

This is an exponential distribution, or the non-normalized χ^2 distribution with two degrees of freedom. If we define the normalized variables

$$x'(t) = \frac{x(t)}{\sigma} \text{ and } y'(t) = \frac{x(t)}{\sigma}$$

then, for $q = [x'^2 + y'^2]$, we then have the χ^2 distribution with two degrees of freedom:

$$p(q) = \frac{1}{2} e^{-\frac{q}{2}}. \tag{B.7}$$

with mean and variance $E\{q\} = 2$ and $V\{q\} = 4$.

B.2 Envelope Squared of a Sinusoid plus Narrowband Noise

Consider the signal model now to be a sine wave with additive Gaussian noise, $n(t)$, so that

$$f(t) = A\cos(\omega_c t + \phi) + n(t),$$

or, equivalently

$$f(t) = [A\cos\phi + x(t)]\cos(\omega_c t) - [A\sin\phi + y(t)]\sin(\omega_c t), \tag{B.8}$$

where ϕ is uniformly distributed in $[0, 2\pi]$. The envelope squared is then

$$u(t) = [A\cos\phi + x(t)]^2 + [A\sin\phi + y(t)]^2. \tag{B.9}$$

The probability density function for u is thus, supressing the time arguments,

$$p(u) = \frac{1}{2\sigma^2} \exp\left[-\frac{1}{2\sigma^2}(u + A)^2\right] I_0\left(\frac{Au^{1/2}}{\sigma^2}\right). \tag{B.10}$$

Equation B.10 is the non-central χ^2 distribution with two degrees of freedom. If we now normalize the variables, as done above in Section B.1, we have a new, normalized non-central χ^2 variable:

$$q = \frac{(A+x)^2}{\sigma^2} + \frac{(A+y)^2}{\sigma^2}. \tag{B.11}$$

And the density for q is a modified form of Equation B.10 as follows:

$$p(q) = \frac{1}{2}\exp\left[-\frac{\gamma+q}{2}\right] I_0((q\gamma)^{1/2}) \tag{B.12}$$

where $\gamma = 2A^2/\sigma^2$ is defined as the non-central parameter for two degrees of freedom.

B.3 DFT Bin Noise Correlation

Theorem:

Given a white Gaussian noise sequence, $n(t)$, such that $E\{n(t)\} = \sigma^2$, and $E\{n(t)n^*(s)\} = \sigma^2\delta_{ts}$, where

$$\delta_{ts} = \begin{cases} 1 & \text{for } t = s \\ 0 & \text{otherwise} \end{cases}$$

is the Kronecker δ function, then $E\{N(k)N^*(l)\} = \sigma^2\delta_{k,l}$ for the DFT sequence $N(k)$ defined as

$$N(k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} e^{\frac{-j2\pi kt}{N}} n(t).$$

Proof:

The autocorrelation function for the DFT sequence $N(k)$ is given by

$$E\{N(k)N^*(l)\} = E\{\frac{1}{N} \sum_{t=0}^{N-1} e^{\frac{-j2\pi kt}{N}} n(t) \sum_{s=0}^{N-1} e^{\frac{j2\pi ls}{N}} n^*(s)\}. \tag{B.13}$$

or

$$E\{N(k)N^*(l)\} = \frac{1}{N} \sum_{t=0}^{N-1} \sum_{s=0}^{N-1} e^{\frac{-j2\pi kt}{N}} e^{\frac{j2\pi ls}{N}} E\{n(t)n^*(s)\} \tag{B.14}$$

where $E\{n(t)n^*(s)\} = \sigma^2\delta_{ts}$ because the process $n(t)$ is white. Then, Equation B.14 is

$$E\{N(k)N^*(l)\} = \frac{\sigma^2}{N} \sum_{t=0}^{N-1} e^{\frac{-j2\pi(k-l)t}{N}} \tag{B.15}$$

This is in the form of a geometric series which we can alternatively express as

$$E\{N(k)N^*(l)\} = \frac{\sigma^2}{N}\frac{1-e^{\frac{-j2\pi bN}{N}}}{1-e^{\frac{-j2\pi b}{N}}} \tag{B.16}$$

where $b = k - l$. Reducing fractions,

$$E\{N(k)N^*(l)\} = \frac{\sigma^2}{N}\frac{1-e^{-j2\pi b}}{1-e^{\frac{-j2\pi b}{N}}} \ . \tag{B.17}$$

The quantity b can only assume integer values $0, 1, \ldots, N-1$. Therefore, from the form of Equation B.17 it is obvious that, since $\exp[-j2\pi b] = 1$ for $b = 1, 2, \ldots, N-1$. the numerator is just $1 - 1 = 0$ and the denominator is non-zero, so

$$E\{N(k)N^*(l)\} = 0 \quad \text{for } k \neq l.$$

However, since the denominator is also zero when $k = l$, L'Hospital's rule must be used to resolve the indeterminate form. This yields

$$E\{N(k)N^*(l)\} = \frac{\sigma^2}{N}\frac{j2\pi be^{-j2\pi b}}{\frac{j2\pi}{N}e^{\frac{-j2\pi b}{N}}} = \sigma^2 \ \text{ for } b = 0 \ (\text{or } k = l.)$$

The autocorrelation for the noise in different bins of the FFT is given then as:

$$E\{N(k)N^*(l)\} = \begin{cases} \sigma^2 & \text{for } k = l \\ 0 & \text{for } k \neq l. \end{cases} \tag{B.18}$$

B.4 Detector Structure

In the following discussion, we will supress all time arguments for the sake of brevity. For the detector design, we will allow two hypotheses for each signal band, or, in our case, each FFT bin. These will be defined for each FFT bin k, as

$$H_{k,0}: F_k = N_k$$

$$H_{k,1}: F_k = A_k + N_k$$

where F_k is the kth bin of the transform of $f(t)$, A_k is the signal magnitude in the kth bin, and N_k is the noise process in the kth transform bin. Since $n(t)$ is a Gaussian process, the noise in the various FFT bins are uncorrelated as shown in Section B.3 in Equation B.18. This allows the use of independent detectors in our application.

The detector then tests on the likelihood of the above two hypotheses and yields a decision $D_k(F)$ as follows:

$$D_k(F) = \begin{cases} 1 & \text{if } P(H_{k,1})p(F_k|H_{k,1}) \geq P(H_{k,0})p(F_k|H_{k,0}) \\ 0 & \text{otherwise} \end{cases} \tag{B.19}$$

where $P(H_{k,i})$ is the *a priori* probability for the ith hypothesis in the kth FFT bin, and $p(F_k|H_{k,i})$ is the hypothesis-conditional probability density function the form of which will be given presently. Equation B.19 can be given as a ratio of the form:

$$D_k(F) = \begin{cases} 1 & \text{if } \frac{P(H_{k,1})p(F_k|H_{k,1})}{P(H_{k,0})p(F_k|H_{k,0})} \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{B.20}$$

or as a log-likelihood function

$$D_k(F) = \begin{cases} 1 & \text{if } \log P(H_{k,1}) + \log p(F_k|H_{k,1}) - \log P(H_{k,0}) - \log p(F_k|H_{k,0}) \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{B.21}$$

The quantities given in Equations B.19 - B.21 will now be defined. The *a priori* probabilities for the binary decision case are given as:

$$\begin{aligned} P(H_{k,0}) &= 1 - \lambda_k \\ P(H_{k,1}) &= \lambda_k. \end{aligned} \tag{B.22}$$

Since the FFT results in a complex output for each bin k, we define some variables which will expedite the application of the χ^2 density functions. Let

$$\begin{aligned} x_{1,k} &= \text{Real}[F_k] \\ x_{2,k} &= \text{Imaginary}[F_k]. \end{aligned}$$

and

$$q_k = \frac{x_{1,k}^2}{\sigma^2} + \frac{x_{2,k}^2}{\sigma^2}$$

And the hypothesis conditional densities are of the forms given in Equations B.7 and B.12. For $H_{k,0}$ we assume that q_k is drawn from a central χ^2 distribution with probability density as follows:

$$p(q_k) = \frac{1}{2} e^{-\frac{q_k}{2}}. \tag{B.23}$$

And for $H_{k,1}$ we assume that q_k is drawn from a non-central χ^2 distribution with probability density as follows:

$$p(q_k) = \frac{1}{2} \exp\left[-\frac{\gamma_k + q_k}{2}\right] I_0((q_k \gamma_k)^{1/2}) \tag{B.24}$$

where

$$\gamma_k = \frac{2A_k^2}{\sigma^2}.$$

When the quantities given in Equations B.23 and B.24 are substituted into Equation B.21, we arrive at the log-likelihood ratio test for this problem presented as

$$D_k(F) = \begin{cases} 1 & \text{if } \log \lambda_k - \frac{\gamma_k}{2} + \log(I_0[(q_k \gamma_k)^{1/2}]) - \log(1 - \lambda_k) \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{B.25}$$

Appendix C

Distribution List

Defense Technical Information Center
Defense Logistics Agency
Cameron Station
Alexandria, VA 22314

Dr. D. Gringras
Code 733
Naval Ocean Systems Center
San Diego, CA 92152

Mr. J. Schuster
PMW-180
Space and Naval Warfare Systems Command
Washington, D.C. 20363-5100

Mr. A. J. Passamante
Code 5032
Naval Air Development Center
Warminster, PA 18974

Dr. N. Owsley
Code 3211
Naval Undersea Systems Center

New London, CT 06320

Dr. N. Gerr

Code 1111

Office of Naval Research

800 N. Quincy Street

Arlington, VA 22217-5000

Dr. T. Warfield

Code 231

Office of Naval Technology

800 N. Quincy Street

Arlington, VA 22217-5000

Mr. N. Smeh

Code 324

Space and Naval Warfare Systems Command

Washington, D.C. 20363-5100

END

DATE

FILMED

5-88

DTIC